Risks of Sharing Cyber Incident Information

Adham Albakri
School of Computing
University of Kent
UK
a.albakri@kent.ac.uk

Eerke Boiten
School of Computer Science and Informatics
De Montfort University
UK
eerke.boiten@dmu.ac.uk

Rogério De Lemos
School of Computing
University of Kent
UK
r.delemos@kent.ac.uk

ABSTRACT

Incident information sharing is being encouraged and mandated as a way of improving overall cyber intelligence and defense, but its take up is slow. Organisations may well be justified in perceiving risks in sharing and disclosing cyber incident information, but they tend to express such worries in broad and vague terms. This paper presents a specific and granular analysis of the risks in cyber incident information sharing, looking in detail at what information may be contained in incident reports and which specific risks are associated with its disclosure. We use the STIX incident model as indicative of the types of information that might be reported. For each data field included, we identify and evaluate the threats associated with its disclosure, including the extent to which it identifies organisations and individuals. The main outcome of this analysis is a detailed understanding of which information in cyber incident reports requires protection, against specific threats with assessed severity. A secondary outcome of the analysis is a set of guidelines for disciplined use of the STIX incident model in order to reduce information security risk.

CCS CONCEPTS

• **Security and privacy → Security services; Human and societal aspects of security and privacy; Governmental regulations.**

KEYWORDS

Information Sharing, Cyber Incident Information, Associated Threats, STIX, Cyber Intelligence.

ARES 2018, August 27–30, 2018, Hamburg, Germany
© 2018 Association for Computing Machinery. ACM ISBN 978-1-4503-6448-5/18/08…$15.00
https://doi.org/10.1145/3230833.3233284

ACM Reference format:

A. Albakri, E. Boiten and R. De Lemos. 2018. Risks of Sharing Cyber Incident Information. In *Proceedings of International Conference on Availability, Reliability and Security, Hamburg, Germany, August 27–30 2018 (ARES 2018), 10 pages.*
DOI: 10.1145/3230833.3233284

1 Introduction

In recent years, cyber attacks have been rapidly increasing in number and complexity [1] [2]. Cyber criminals are becoming more organized, and cyber attacks are perpetrated for many reasons. The goal may be to obtain personal and sensitive information, such as credit card details or medical information, extortion, or to disrupt computer systems and critical infrastructures.

There is a need for awareness and mitigation of threats even before the criminals start their attacks, particularly where critical assets are concerned. Understanding of previous attacks can provide the necessary information for protecting organisations. Cyber intelligence information sharing involves the collection and analysis of cyber threat information. This can be done between peer organisations in the same sector (government, telecom, banks, or health), or through central authorities such as cybersecurity incident reporting teams (CSIRT). An effective incident reporting procedure using this information will increase the ability to provide timely responses to incidents, including alerts to information sharing parties.

In order to assist the sharing process of cybersecurity information many standards and platforms are available such as STIX [3] and CybOX [4]. STIX is a language developed to represent and standardize cyber threat information. We will use the STIX incident model as an indicative representation of what may be included in cyber incident information.

There are several legal factors that influence when and how cyber information sharing can take place. The EU's new General Data Protection Regulation (GDPR) [5] will be a driving motivation for organisations to maintain high quality cyber security. In support of that, it also explicitly allows proportionate processing of personal data in order to support its overall goal of keeping personal information secure. The GDPR and the Directive on Security of Network and Information Systems (NIS Directive) [5] both mandate cyber incident reporting to authorities, and the latter encourages sharing for increased resilience. Many guidelines and studies highlight the presence of sensitive information elements within cyber threat information. In particular, NIST guidelines for cyber threat information sharing include rules for establishing information sharing relations [6]. They introduce examples of sensitive and identifiable information that may be present when sharing different types of threat information. However, they do not analyse the threats that arise as a consequence of sharing sensitive and identifiable information. In this paper we provide the first detailed analysis of this.

Our overall goal is to improve and stimulate cyber information sharing, while mitigating its potential adverse effects. In this paper we provide a detailed understanding of the threats arising from sharing. Our systematic analysis of the threats in cyber incident information proceeds through the explicit consideration of the risks posed by the various pieces of information that could be included and the associated threats that they enable. By doing so, we show the risk of disclosing any information in the STIX incident model. The analysis also highlights areas where the necessarily flexible use of STIX induces additional disclosure risks; this naturally leads to guidelines on the use of STIX in order to mitigate these.

The remainder of this paper is organized as follows. Section 2 provides background regarding cyber threat information sharing and its legal basis. Section 3 describes the methods used for threat analysis. Section 4 discusses the analysis of disclosing cybersecurity incident information in the STIX incident model with its the key findings. Finally, Section 5 briefly concludes the paper and proposes future research directions.

2 Background

This section provides general background, discussing models of cyber threat information sharing and in particular STIX, as well as the broader landscape of obstacles to cyber incident sharing and the legal context.

2.1 Cyber intelligence sharing methods

Cyber intelligence sharing covers incident reports as well as other types of information such as threats, vulnerabilities, mitigations, situational awareness, best practices and strategic analysis. Taking into consideration the breadth and quantity of information that might be exchanged, structured methods and automated systems are essential to make this practicable. Security software vendors' websites and "white papers" frequently present solutions, but it cannot entirely be disregarded that their main motivation is to sell security solutions. Standards organisations and researchers have started to develop and provide models and systems such as "Threat Intelligence Sharing Platforms" [7]. These platforms provide automated support to information sharing and associated analysis. Several standards have been proposed, and others are still under development, for the automated exchange of cyber threat information. These include Cyber Observable eXpression (CybOX™) [4], Structured Threat Information Expression (STIX™) [8], An Open Framework for Sharing Threat Intelligence (OpenIOC), Incident Object Description Exchange Format (IODEF) [9] and Automated eXchange of Indicator Information (TAXII) [10][7]. In this paper, we will be using STIX, described next.

2.2 Structured Threat Information eXpression (STIX)

Structured Threat Information eXpression (STIX) [3] is a language for representing cyber threat information. It was developed in collaboration under the OASIS umbrella by a variety of parties interested in specification, capture, characterization and communication of standardized cyber threat information. STIX provides an architecture to support several components used to express the core of threat concepts, including: Cyber Observables, Indicators, Incidents, Adversary (Tactics, Techniques, and Procedures) (TTPs), Exploit targets, Courses of action, Cyber Attack Campaigns and Threat actors. STIX is considered the most commonly used standard in commercial products to automate information sharing [7]. To share STIX reports, we can use a standard called Trusted Automated Exchange of Intelligence Information (TAXII), which is an application layer protocol used to exchange cyber threat information in STIX over HTTPS [10].

2.3 Perceived threats and challenges for information sharing

Organisations perceive significant barriers and challenges to sharing information for cyber intelligence analysis. The most obvious of these is the risk associated with disclosing sensitive information. Other barriers include trust among users and between users and platforms providers, different privacy laws, as well as technical issues arising from different platforms and standards [11] [12].

It is necessary to establish a win-win environment where all entities get the benefit from sharing information, and avoiding entities that do not cooperate or just want to get benefit from the others ("free-riders"). In general, trust between the sharing partners needs to be established. One simple method of achieving this is to share information via a trusted central authority such as CERT-UK or CISP in the UK. Industry sector regulators could also be considered for this, but the punitive dimension of regulation may be a factor that inhibits the sharing of information.

The largest perceived threats arise from information disclosure risks. Shared information about incidents may contain: sensitive information related to the impact of the incident, the affected assets, personal information and data belonging to the victims and the incident reporters, information about the organization's cybersecurity strengths and weaknesses, as well as competition-sensitive information about business processes. Threats may be to the organisation's reputation or derived from concerns of intellectual property, business confidentiality or data protection. However, this risk has not been previously analysed in detail.

2.4 The legal context

Cyber information sharing takes place in a legal context. Laws and regulation may both encourage and inhibit aspects of cyber information sharing. The main relevant laws for this in the EU context are the General Data Protection Regulation (GDPR), and the Directive on Security of Network and Information Systems (NIS Directive), both having come into force in May 2018.

GDPR. The GDPR is the primary law that sets out requirements for any companies processing personal data of EU citizens or from within the EU. As an EU Regulation it is immediately binding to all member states [5][13].

The GDPR is concerned with the protection of "personal data", which is "any information relating to an identified or identifiable natural person [...]". In particular, this may also include "online identifiers", such as IP addresses in certain contexts. There are also "special categories" of personal data, called "sensitive personal data" in previous laws; these will be infrequent in cyber incident reports, and we will use the term "sensitive information" in this paper in an informal rather than this legal sense.

The GDPR is expected to be a main driver for improving cyber security in Europe in the near future, as it asks to "implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk" (Art.32). Such measures might include intelligence sharing – but what if there is personal data contained in that? Figure 1 shows some of the STIX incident model properties which are more and less likely to contain personal data under the GDPR.

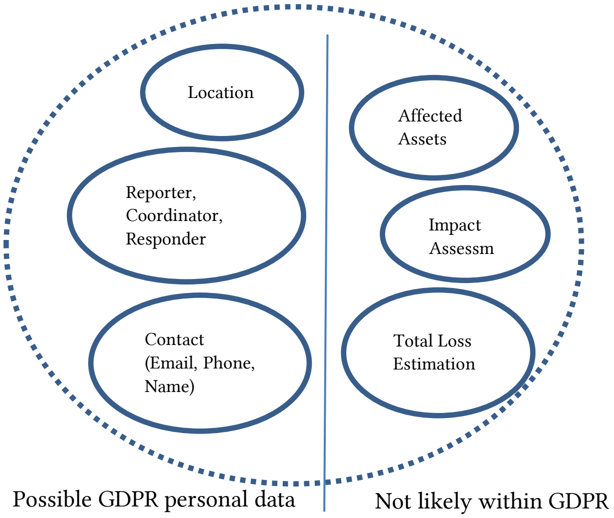
Location
Reporter, Coordinator, Responder
Contact (Email, Phone, Name)
Affected Assets
Impact Assessm
Total Loss Estimation
Possible GDPR personal data
Not likely within GDPR

Figure 1 Examples of the STIX incident properties

Any processing of personal data needs to be done on a legal basis, with Art.6(1) listing five possible bases besides consent, of which "legitimate interest" is of particular interest. This is because Recital 49 of the GDPR admits a legitimate interest for "processing for the purpose of ensuring network and information security, including preventing unauthorised access to electronic communications networks, and stopping damage to computer and electronic communication systems". For any processing such as this which is based on legitimate interest, this justification needs to be balanced with potential adverse impact on the data subjects. In addition, one might consider the mandatory breach notifications of Art.33 to the relevant supervisory authority as a form of cyber incident information sharing.

NIS Directive. The NIS Directive is the EU's first piece of cyber security legislation. It requires the establishment of "Competent Authorities" (CA) which represent and regulate cyber security within critical sectors in their countries. Operators of Essential Services (OES) and Digital Service Providers (DSP) are required to report any incident affecting the availability, authenticity, integrity or confidentiality of data transmitted, stored or processed to the relevant CA. The CA in different countries are expected to share cyber intelligence. In addition, the CA can audit the OES/DSP's cyber security provision, as well as instruct them on improvements. Each EU member state will have needed to define legislation to implement the Directive, including methods for determining which organisations are OES/DSP as well as "effective, proportionate and dissuasive" penalties for infringement [14]. The NIS Directive assigns the European Union Agency for Network and Information Security (ENISA)[15] a central role in providing cybersecurity advice and solutions.

3 Threat analysis methods

In our analysis of the data fields in the STIX cyber incident model, we will be indicating the roles the various attributes may play: they could contain sensitive information, or help to identify people and organisations. This is explored in Section 3.1 below. We then point out threats, using a taxonomy briefly described in Section 3.2. Finally, we assess the severity of privacy and security threats according to a methodology described in Section 3.3.

3.1 Sensitive information and the identification categories of attributes

For categorizing sensitivity of data items in cyber incident reports, we use common characterizations from the literature on anonymization and de-identification methods [16] [17] . The attributes' types are [18]:

Identifier attributes include information used to identify an individual such as full name, driver license, and social security number.

Quasi-identifier attributes include attributes that can be used together, or linked with an external source, to re-identify individuals, such as gender, age, date of birth, postcode.

Sensitive attributes include information that should be confidential, examples include disease, salary, etc.

Insensitive attributes are all other attributes and innocuous information.

For a disclosure to be harmful, it needs to contain *sensitive* information about an *identifiable* subject. Although some attributes are not sensitive or identifying by themselves, combining them with other attributes may reveal sensitive information and identify organisations and individuals.

3.2 Threat taxonomy

In our systematic analysis that follows, we will use the threat taxonomy from ENISA [19] for categorizing the threats. The high level categories of this taxonomy are:

- Physical attack (deliberate/ intentional)
- Disaster (natural, environmental)
- Failures/ Malfunction
- Outages
- Eavesdropping/ Interception/ Hijacking
- Nefarious Activity/ Abuse
- Legal
- Unintentional damage/loss of information or IT assets

3.3 Severity analysis of threats

In traditional risk assessment, risks are evaluated for *impact* and *likelihood*. The latter is particularly problematic for risks that require action by an attacker to materialise: we would need to find out how likely it is that some attacker will be motivated to exploit a given weakness. To avoid having to guess that motivation, we assess *exposure*: how easy would it be for a motivated attacker to exploit, and what prejudicial effects might be caused? This approach is taken for privacy risk in the standard for privacy risk management by the French data protection authority CNIL (Commission Nationale de l'Informatique et des Libertés) [20]. We have generalized this to apply to cyber security risks as well. For privacy risks, the exploitability depends on how easy it would be to identify a specific individual, i.e. the level of identification. Table 1 shows the description of the scores for this on a 1-4 scale, as taken from [20] .

Table 1 Level of Identification

Score	Meaning	Description
1	Negligible	Impossible to identify the individual
2	Limited	Possible but difficult to identify the individual
3	Significant	Relatively easy to identify the individual
4	Maximum	Extremely easy to identify the individual

The **prejudicial effects** value of each threat is also scored on a 1-4 scale as given in [20]. Table 2 describes this.

Table 2 Prejudicial Effects

Score	Meaning	Description
1	Negligible	There is no problem
2	Limited	It could be inconvenient to the individual, partially affecting the system
3	Significant	There are significant consequences, with serious difficulties
4	Maximum	There are critical irrevocable consequences

Finally, The CNIL standard [20] computes the severity value by adding the level of identifiability and prejudicial effects of potential impacts values obtained and translates that into a risk severity scale as given in Table 3. We record the resulting severity as PS (Privacy Severity) for each risk in our analysis.

Table 3 Severity Value

Level of identification + Prejudicial effects	Corresponding Severity
< 5	1. Negligible
= 5	2. Limited
= 6	3. Significant
> 6	4. Maximum

As indicated above, we have generalized this method to also apply to cyber security risks, yielding Cybersecurity Severity (CSS) score. For this, we use Table 1 and Table 3 unchanged, and instead of Table 2 we use the very similarly constructed Table 4 to score the ease of exploiting cybersecurity information.

Table 4 Ease of Exploitation

Scor e	Meaning	Description
1	Negligible	Impossible to exploit cybersecurity information
2	Limited	Possible but difficult to exploit cybersecurity information
3	Significant	Relatively easy to exploit cybersecurity information
4	Maximum	Extremely easy to exploit cybersecurity information

4 Information disclosure threat analysis of the STIX incident model

In the following, we apply the methods described above to the STIX incident model. We illustrate what are the threats associated when disclosing any particular property in the incident model and identify the level of sensitivity and identification, as well as the severity of any associated threats.

4.1 Information recorded in the analysis

The overall objective of our analysis is to establish which information in cyber incident report needs to be protected and why. In order to achieve that, we take the STIX incident model as indicative for what might be included in such reports. For each property in every class of the STIX incident model we assign the threats associated with its disclosure based on its sensitivity and identification level. We analyse a total of 123 properties. Table 5 shows the analysis of an illustrative subset of attributes. Each STIX property is recorded in one row in the table, with labelled columns representing the relevant analysis and description. The columns: Complex Type, Include Free Text, Sensitivity, Identification, Personal information, Justification, Threat, Privacy Severity and Cybersecurity Severity contain our analysis of these properties.

The *Complex Type* column indicates that the property's type is a composite of other types. Therefore, its analysis may be derived from that of the component types.

The *Include Free Text* column indicates that the property or one of its constituents is a free text field. In principle, any information could be exposed through such an unconstrained field. Taking this to an extreme would trivialize our analysis: most of the information contained in an incident report would be potentially sensitive and identifying. We take a pragmatic approach to this: our analysis is based on the assumption that the person who is responsible for filling in the report will insert only information consistent with the property description and the context of the report. We acknowledge a vulnerability in the STIX incident model regarding information leaks, here and in general, due to the lack of constraints on fields. Minimizing the impact of this on information security requires a disciplined use of the model, as discussed later.

The *Sensitivity* column indicates whether the property includes information that presents a confidentiality risk, such as IP addresses or the assets affected in the incident. In our analysis, we give for each property a sensitivity value, which will be either “Yes”, “No”, or “It depends”:

Yes: includes information that should be confidential, for example, financial information and the vulnerability exploited in the incident.

It depends: not necessarily sensitive but it could be in some cases; the *Justification* column then contains further elaboration of the circumstances.

The *Identification* column indicates whether the property could identify an individual or the organisation. For each property, we provide an identification value, which will be one of the following:

- Yes: it is information that likely identifies an organisation or an individual.
- No: knowing this information will not be helpful in identifying an organisation or an individual.
- Quasi Identifier (QI): the information could be linked with other information or an external source to re-identify an individual or the organisation.

For identifying personal information that refers to individuals rather than organisations, we have added a *Personal information* column to indicate that the disclosure of the property could reveal personal information.

The *Threat* column indicates the possible threats when revealing information associated with the property, based on the property description, sub-properties in case it is complex, and the actual information.

The severity of the threats is given in the *PS* and *CSS* columns with scores assigned as described in Section 3.3. In fact in the table we include the original scores as e.g. 2+3 for exploitability and impact without translating to the 1-4 scale as per Table 3. The goal of this exercise is to identify potential threats when sharing incident information, to provide an explanation what the sensitivity and identifiability are, and ultimately to address the potential threat when disclosing information associated with properties of the STIX incident model.

4.2 Analysis sample

Table 5 gives an example of some properties in the IncidentType class of the STIX incident model. Cells in columns "Complex Type" (CT), "Include Free Text" (IFT) , "Sensitivity" (S), "Identification" (I), "Personal Information" (PI), "Justification" and "Threat" represent our analysis. The values in the column "Property" are summarized from the STIX incident model.

This table gives grounds behind our analysis of properties. Some properties have only a cybersecurity severity value such as "Security_Compromise", and some properties have both privacy and cybersecurity severity values, such as "COA_Requested", which contains identifiable information for the source of information, in addition to the sensitive information about the system and the infrastructure as well. We explain the values of PS and CCS for the following properties:

"Description" property: It is a free text field to describe the incident. It is not unlikely that the reporter will include critical information in this field, which could contain cybersecurity and identifiable information. The PS value is 2+2, as the level of identification is 2: it is possible to identify individuals with difficulty. The second value is the prejudicial effect which is also 2 due to the possible disclosure of the identity without further information. Similarly, the CSS is scored as 2+2 by assigning 2 as the difficulty of exploitation (any vulnerabilities are likely described at a very high level in this field), and 2 as the prejudical effects due to the problem of the data breach.

"Reporter" property: this contains both privacy and cybersecurity threats. Since it contains explicit information about the reporter, it is very easy to identify the person. One of the possible outcomes might be identity theft. The cyber security risk is in revealing the identity of what is likely a good target for a spear phishing or other social engineering attack.

"Security_Compromise" property: This property does not contain any identifiable information therefore the privacy severity is zero. On the other hand, the CSS is 3+3: 3 related to how easy to exploit which security sector has been compromised and based on that there is a possibility to perform many types of cyber-attacks such as an integrity attack on the data. For example, it could be a backdoor attack based on disclosed vulnerability that gives remote access to the victim's system.

"COA_Requested" and "Related_Indicators" properties both contain a privacy threat because of the "InformationSourceType" property in it, which might identify an individual. They also include a substantive cyber threat due to the course of action that implies vulnerabilities and the technical information such as IP addresses and information about network traffic that might be revealed.

4.3 Severity results

We have computed the severity values for each property of the STIX incident model based on the method proposed in Section 3.3. In Particular, Figure 2 shows the cybersecurity severity results for the first level properties of the STIX incident model.

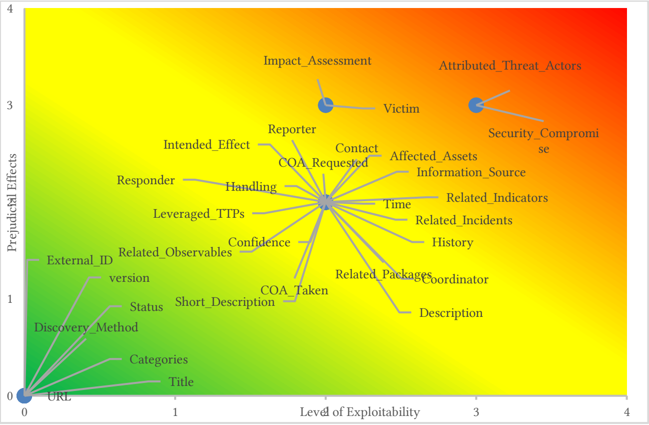

Figure 2 Cybersecurity Severity Results

Table 5 Analysis Sample for Class “IncidentType”

Property	CT	IFT	S	I	PI	Justification	Threat	PS	CSS
version	✗	✗	✗	✗	✗	This refers to the report, not to the incident	N/A	0	0
Description	✓	✓	✓	✓	*	free text field which is likely to refer to particular business information, and may contain sensitive and identifying information	Compromising confidential information (data breaches); Social Engineering; Violation of laws or regulations / Breach of legislation.	2+2	2+2
Reporter	✓	✓	✗	✓	✓	The identity of the reporter can be revealed	Identity theft (Identity Fraud/ Account); Social Engineering; Unauthorized activities	4+2	2+2
Security_ Compromise	✗	✗	✓	✗	✗	Identifies whether critical information was leaked, and it can be sensitive	loss of reputation; Loss of (integrity of) sensitive information	0	3+3
COA_ Requested	✓	✓	✓	✓	✓	This can refer to specific information about the business, and how the organisation can return to business as usual. Recovery operation includes its own security risks which may be exploited in a targeted attack.	Man in the middle; Social Engineering; Generation and use of rogue certificates; Compromising confidential information (data breaches); Targeted attacks (APTs etc.)	2+2	2+2
Related_ Indicators	✓	✓	✓	✓	✓	This can refer to specific information about the incident and adversary Tactics, Techniques, and Procedures (TTPs). May contain identifying information about the adversary	Loss of (integrity of) sensitive information; Social Engineering; Misuse of information/ information systems; Unauthorized activities Compromising confidential information (data breaches); Failure to meet contractual requirements	2+2	2+2

Analysis Sample for Class “*IncidentType*”. In columns, (CT) stands for Complex Type, (IFT) for Include-Free-Text, (S) for Sensitivity, (I) for Identification, (PI) for Personal Information, (PS) for Privacy Severity, (CSS) Cybersecurity Severity. For the values of the properties, ‘✓’ denotes ‘yes’, ‘✗’ denotes ‘No’, ‘*’ denotes ‘It depends’.

Figure 3 shows the privacy severity results for the first level properties of the STIX incident model.

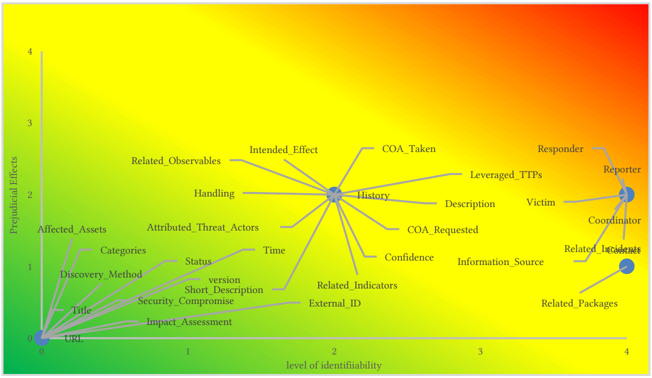

Figure 3 Privacy Severity Results

At first glance it may be surprising that Prejudicial Effects never achieve the highest score 4, for irrecoverable damage. Our explanation for this is rather different between the two dimensions. In the privacy dimension, this is an impact of the particular context of cyber incident reporting. Personal data never plays a central role in this, and there is no sensitive personal data involved in this scenario at all. Thus, any privacy risks will be limited.

For the cyber security dimension, it is due to the nature of cyber security itself. It is extremely rare for a successful cyber attack, particularly in a critical infrastructure context, to exploit only a single vulnerability. Conversely, exploiting a single vulnerability is always unlikely to lead to irrecoverable damage by itself.

This suggests an extension to our analysis per property is necessary. For a full awareness of overall risks, we need to look at *combinations* of properties that together provide a feasible composite attack threat. Although this is in theory unfeasible (nearly 2^{123} combinations of the 123 different properties), it can be triaged by focusing on known effective combinations of types of threats and the most severe individual threats.

As an illustration, we describe a composite threat that could lead to irrecoverable damage to the system. In order to launch any serious attacks, the attackers need to collect data about the target’s activity. The ‘Reporter’ property will be an entry point for online research leading to a social engineering attack. This may lead to the installation of a key logger or other malware. The ‘Security_Compromise’ property might then reveal which security hole in a critical system can be exploited starting from the Reporter’s computer. A real-world example of a successful attack against critical infrastructure is the Ukraine Attack [21]. This attack started by weaponising the network with BlackEnergy malware using spear-phishing attacks, then hijacking SCADA systems, and remotely controlling electricity substations.

4.4 Key findings

The analysis has provided a broad and detailed insight into the disclosure risks associated with cyber incident reports, when encoded in the STIX incident model. It has highlighted individual pieces of sensitive information as well as the specific threats arising from their disclosure.

The STIX incident model consists of a hierarchy of classes containing 123 properties, and these were analysed separately. Properties may be sensitive both through their immediate content and through their specific context within complex properties. For example, the "Reporter" property tells us not only an employee name but also identifies the person who had reported the incident and so is likely in a central cybersecurity role in the organisation. The object-oriented structure of the STIX incident model implies that some sensitivity arises also through class inheritance: it may be inherited from a superclass, as well as arise in a specific subclass. In the following, we present general observations that follow from the analysis performed on the STIX incident model.

Controlled/Uncontrolled properties identified in STIX incident model. STIX is designed to be flexible and liberal about the information contained and how it is represented. The incident model suggests specific value sets for many properties, but also allows the content creator to choose any arbitrary value. This lack of constraints implies that undisciplined use may disclose arbitrary sensitive information. In particular, many properties consist of free text, which may contain critical information about the incident, including organisation name, IP addresses, impact and Course of Action, that must be protected. Tools for extracting sensitive and identifying information from text are available: these can be characterized as rule-based or machine learning-based [22]. The rule-based tools usually handle the re-identification goal with pattern matching, regular expressions and dictionary lookups. For example, the strings "DDoS" and "146.227.156.60" within some free text property could be classified into the categories of incident category and IP addresses.

Categories of information and associated threats. Intuitively, we expected to find threats relating to different kinds of information disclosure: personal, organizational, financial and cybersecurity. Indeed, most STIX properties related specifically with one of these kinds, and have a matching set of associated threats. Moreover, for each of these types a significant number of properties is present in the STIX incident model.

Disclosing personal information. The number of properties that identify individuals in the organisations is high, such as the Reporter property that characterizes the entity that reported the incident, and the Responder property that characterizes the entity playing the role of the responder for the Incident. Thus, disclosing any of these properties will be associated with multiple threats including targeted attacks (APTs etc.) and social engineering attacks, such as phishing and spear phishing. In [23], CERT-UK provides a case study of targeting a system administrator of a UK organization by a spear phishing attack. The attackers identified the system administrator and sent a spam email to the system administrator. The goal of this attack was to install a RAT (Remote Access Trojan) and getting advantage of the administrator permission to get access to the network and collect sensitive information about the critical systems in this targeted organization.

Disclosing the organisation's information. The number of properties that potentially identify organisations is high. For example, the *Affected_Asset* property that specifies a list of one or more assets affected includes a description of the asset and the security effect on the asset, for example, a HR database server for an organisation. Thus, disclosing any of these properties will be associated with threats including physical attack as well as targeted attacks and social engineering.

Disclosing financial information. The STIX incident model contains specific financial information that covers the estimated cost to the victim, which is based on the loss of revenue from system downtime and operation cost to fix the damage. For example,
the *Total_Loss_Estimation* property specifies the total estimated financial loss for the Incident and the *Response_And_Recovery_Costs* property specifies the level of response and recovery-related costs. The loss of this confidential information forms a data breach threat by itself but it also has an associated threat of loss of reputation.

Disclosing cybersecurity information. The STIX incident model contains cybersecurity information about the incident, such as the *Course_Of_Action* property. This property refers to the course of action requested and taken for the incident. In addition, it includes specific information about the incident, such as whether non-public data was compromised and whether that data was encrypted or not. The organisation's analysis of the incident can be reported through the *Leveraged_TTPs* property**.** Tactics, Techniques and Procedures (TTPs) consists of the specific adversary behavior (attack patterns, malware, exploits) exhibited and resources leveraged (tools, infrastructure, personas) [24]. This information contributes to providing a complete understanding of the magnitude of the threat. However, disclosing cyber information details like these could give hackers a road map to conducting additional targeted attacks including physical ones.

Some information is critical only in combination. Some properties are in general not sensitive, but become critical when combined with other properties or externally available information. For example, the *First_Malicious_Action* property specifies the time that the first malicious action related to the Incident occurred. This information is not sensitive by itself, but patterns in this information may lead to attribution (identification of the attacker) [25]. As an extreme example, for financial damages, neither the Amount nor the *Iso_currency_code* property by itself is sensitive; however, together they specify the estimated financial loss, which is sensitive. We have discussed the issue of critical combinations of cyber security vulnerabilities in detail in Section 4.3.

4.5 The use of the STIX incident model

As our analysis above indicates, there are clear drawbacks to the flexibility of the current STIX incident model. From the perspective of disclosure, free text fields and unconstrained properties allow for information leaks. In addition, they offer little perspective for data validation and thus scope for undetected human errors. The potential for automated processing is also greatly reduced by variability of inputs. This calls for disciplined use of the STIX model, which is likely most easily provided by ensuring that the more flexible fields are filled through templates, possibly by a system generating STIX reports for the user from higher level information. (As STIX is XML based, which is not intended for human reading and writing, some such interface is essential for human interaction in any case.) Sector organisations could also develop custom versions of the STIX incident model that specialize to their specific risk profile. Implementation of STIX in cyber information sharing platforms could actively support this. In any case, consistent and disciplined use of incident reporting should be supported by appropriate training and policies within individual organisations.

5 Conclusion and future work

We have performed a comprehensive analysis of incident reporting information through the STIX incident model to identify the threats of disclosing sensitive and identifying information. We assigned the sets of possible threats based on the ENISA threat taxonomy. We identified the threats associated with each property, and evaluated those for severity in both the privacy and cyber security dimension. We now have a full overview of which incident information needs protecting, and why. In addition, we have provided guidance for disciplined use of the STIX incident model to reduce and focus information security risks.

Our overall goal is to improve and stimulate cyber information sharing, while mitigating the potential adverse effects. The risk analysis of information sharing should provide organisations the means for making evidenced decisions on what information to share, and with whom. More sophisticated methods of sharing use privacy preserving techniques to reduce exposure risks [26]. Applicability of such techniques depends not only on the information, its sensitivity, and the level of trust in the data sharing partner, but also on the analysis to be performed. Ultimately, the sharing choices need to balance preserving confidentiality with preserving utility of the analysis. Thus, an exploration of such analysis operations is the next item on our agenda.

ACKNOWLEDGMENT

This work has received funding from the European Union Framework Programme for Research and Innovation Horizon 2020 under grant agreement No 675320.

REFERENCES

[1] ENISA, "ENISA Threat Landscape Report 2016: 15 Top Cyber-Threats And Trends," 2017. [Online]. Available: https://goo.gl/N3xP1F. [Accessed: 25-Apr-2018].

[2] PricewaterhouseCoopers, "Global State of Information security Survey," 2014.

[3] S. Barnum, "Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™)," *MITRE Corp. July*, pp. 1–20, 2014.

[4] MITRE, "Cyber Observable eXpression," 2011. [Online]. Available: http://cyboxproject.github.io/. [Accessed: 16-Jul-2017].

[5] European Parliament Council of the European Union,

“DIRECTIVE (EU) 2016/1148 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 6 July 2016,” *Off. J. Eur. Union*, 2016.

[6] C. S. Johnson, M. L. Badger, D. A. Waltermire, J. Snyder, and C. Skorupka, “Guide to Cyber Threat Information Sharing,” 2016. [Online]. Available: https://goo.gl/8dBBxE. [Accessed: 25-Apr-2018].

[7] C. Sauerwein, C. Sillaber, A. Mussmann, and R. Breu, “Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives,” in *13th International Conference on Wirtschaftsinformatik (WI 2017)*, 2017, pp. 837–851.

[8] MITRE, “Structured Threat Information eXpression,” 2014. [Online]. Available: http://stixproject.github.io/. [Accessed: 16-Jul-2017].

[9] Y. Danyliw, R Meijer, J Demchenko, “The Incident Object Description Exchange Format,” 2007. [Online]. Available: https://www.ietf.org/rfc/rfc5070.txt. [Accessed: 16-Jul-2017].

[10] J. Connolly, M. Davidson, and C. Schmidt, “The Trusted Automated eXchange of Indicator Information (TAXII ™),” 2014. [Online]. Available: https://goo.gl/GqbVfK. [Accessed: 25-Apr-2018].

[11] B. Petrenj, E. Lettieri, and P. Trucco, “Information sharing and collaboration for critical infrastructure resilience - a comprehensive review on barriers and emerging capabilities,” *Int. J. Crit. Infrastructures*, vol. 9, no. 4, p. 304, 2013.

[12] A. K. Eric Luiijf, “Sharing Cyber Security Information,” 2015. [Online]. Available: https://goo.gl/afbo7J. [Accessed: 25-May-2018].

[13] L. Kalman, “The GDPR and NIS Directive A new age of accountability , security and trust ?,” 2017. [Online]. Available: https://goo.gl/ky3ju4. [Accessed: 25-May-2018].

[14] ENISA, “Incident notification for DSPs in the context of the NIS Directive,” 2016. [Online]. Available: https://goo.gl/FCnuFj. [Accessed: 25-May-2018].

[15] ENISA, “European Union Agency for Network and Information Security,” 2004. [Online]. Available: https://www.enisa.europa.eu/. [Accessed: 22-Jul-2017].

[16] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, “ℓ-Diversity: Privacy beyond k-anonymity,” in *Proceedings - International Conference on Data Engineering*, 2006, p. 24.

[17] L. Ninghui, L. Tiancheng, and S. Venkatasubramanian, “t-Closeness: Privacy beyond k-anonymity and ℓ-diversity,” *Proc. - Int. Conf. Data Eng.*, no. 2, pp. 106–115, 2007.

[18] K. Naganuma, M. Yoshino, H. Sato, and Y. Sato, “Privacy-preserving analysis technique for secure, cloud-based big data analytics,” *Hitachi Rev.*, vol. 63, no. 9, pp. 577–583, 2014.

[19] ENISA, “Threat taxonomy: A tool for structuring threat information. Initial report.,” 2016. [Online]. Available: https://goo.gl/YugeQu. [Accessed: 25-May-2018].

[20] Commission Nationale de l’Informatique et des Libertés, “Methodology for Privacy Risk Management; How to implement the Data Protection Act,” 2012. [Online]. Available: https://goo.gl/o3aN85. [Accessed: 25-Apr-2018].

[21] M. McElfresh, “Cyberattack on Ukraine grid: here’s how it worked and perhaps why it was done,” 2016. [Online]. Available: https://goo.gl/k1KNQP. [Accessed: 22-May-2018].

[22] S. M. Meystre, F. J. Friedlin, B. R. South, S. Shen, and M. H. Samore, “Automatic de-identification of textual documents in the electronic health record: a review of recent research.,” *BMC Med. Res. Methodol.*, vol. 10, no. 1, p. 70, 2010.

[23] CESG, “Common Cyber Attacks: Reducing The Impact,” *UK Government*, 2015. [Online]. Available: https://goo.gl/sdozWr. [Accessed: 25-Apr-2018].

[24] MITRE, “TTPType.” [Online]. Available: https://stixproject.github.io/data-model/1.2/ttp/TTPType/. [Accessed: 06-Dec-2017].

[25] U.S. District Court, “US Dept. of Justice Indictment Chinese Hack,” 2014. [Online]. Available: https://www.justice.gov/iso/opa/resources/5122014519132358461949.pdf. [Accessed: 16-Nov-2017].

[26] J. M. de Fuentes, L. González-Manzano, J. Tapiador, and P. Peris-Lopez, “PRACIS: Privacy-preserving and aggregatable cybersecurity information sharing,” *Comput. Secur.*, vol. 69, pp. 127–141, 2017.