MANAGING CYBERSECURITY AND PRIVACY RISKS OF CYBER THREAT INTELLIGENCE

By

Adham Albakri

March 2021

A thesis submitted to
The University of Kent, School of Computing
in the subject of Computer Science for the degree of Doctor of Philosophy.

Supervised by:
Prof. Eerke Boiten
Prof. Peter Rodgers

Declaration

I declare that the work was solely conducted during registration for the above award with the University of Kent, under University supervision. I declare that no material contained in the thesis has been used in any other submission for an academic award.
I confirm that the work represented in this submission was undertaken solely by myself except where otherwise attributed.

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

I have been actively participating and collaborating in "Air4ICS: Agile Incident Response For Industrial Control Systems" project. The AIR4ICS project is funded by NCSC/RITICS.

Acknowldgment

The journey is easier and more enjoyable when you travel with a good companion. This thesis is the result of four years of hard work during the uncertain times and COVID-19 pandemic, where I have been encouraged and supported by many obliging people.

First, I would like to express my deepest gratitude to my supervisor, Prof Eerke Boiten, for his precious observations and continuous support since the early stage of this research. I am very blessed to have such a dedicated supervisor.

Second, my sincere thanks to Prof Peter Rodgers for his valuable comments and suggestions. Also, I would like to thank Dr Rogério de Lemos for his supervision and feedback during the first 1.5 years of my research.

I am highly grateful for the grant from the NeCS project, a network funded by the European Union Horizon2020 Marie Skłodowska-Curie Actions program. This life-changing opportunity has contributed to my research, training, and networking.

I would also like to express my gratitude to my doctoral committee for their encouragement and invaluable comments. Furthermore, I would like to thank the school of computing at the University of Kent and the cyber technology institute at De Montfort University, where I have spent most of my research as a research visitor.

My appreciation also goes out to my family for their support throughout these years. I am grateful especially to my parents for their infinite care. I want to commend the support I have got from my fellow researchers and friends. Also, my special thanks to my friends who are my main supporters, especially during my PhD; Imad, Dr Ali, M Khattab, Alva, Fenia and many others who were always there during this exciting journey.

Abstract

In recent years, the number of cyber-attacks that affect critical infrastructures such as health, telecommunications and banks has been rapidly increasing. Sharing Cyber Threat Intelligence (CTI) is being encouraged and mandated as a way of improving overall cyber intelligence and defence, 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. There are risks of breaching regulations and laws regarding privacy. With laws and regulations such as the General Data Protection Regulation (GDPR), the managers of CTI datasets need clear guidance on how and when it is legal to share such information. This thesis supports the decision of sharing CTI datasets as it proposes a novel contribution through a detailed understanding of which information in cyber incident reports requires protection against specific threats with assessed severity.

It 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. It provides a set of guidelines for the disciplined use of the STIX incident model in order to reduce information security risk. Then, it proposes a quantitative risk model to assess the risk of sharing CTI datasets enabled by sharing with different entities in various situations. The evaluation of the cyber incident model analysis and the quantative risk model has been validated by means of experts' opinions.

As a final contribution, this thesis defines the impact that GDPR legal aspects may have on the sharing of CTI that helps technical people and CTI managers with limited legal expertise to encompass legal consideration before sharing CTI datasets. In addition, it recommends protection levels for sharing CTI to ensure compliance with the GDPR.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Over the past three decades, the internet has become a crucial part of the way we live and work. It has entered every sector and how we communicate, store, transfer and process information. At the same time, communities, businesses, and governments rely on these technologies. The digital economy is expanding rapidly, driven by generating, analysing and collecting information. This digital information grows from individuals' digital footprints, business work streams, evolving internet of things (IoT) and more online activities. This growth creates many new roles and opportunities; for example, according to 2017 and 2019 digital economy reports by the United Nations, more than 100 million individuals work in the ICT sector. In 2015, e-commerce sales were about $25.3 trillion [4]. By 2030, general-purpose technology such as data analytics is expected to provide an additional economic output of around $13 trillion [5]. Everyone is using a smartphone, computer and IoT devices that connect to the internet. We store personal and confidential information and use it for online banking, shopping and communication via emails and social media. At the same time, cybersecurity threats are evolving. Therefore, we need to exercise caution and define countermeasures to protect the confidentiality, integrity, and availability of the systems and services. It is difficult to come up with a specific definition of cybersecurity and what it contains. Cyber attacks are becoming more sophisticated and creative in different ways. Thus, cybersecurity requires attention and commitment. Currently, it is not difficult for any user to obtain malware and try to perform a cyberattack against any organisation. Cybersecurity protects the devices connected to an organisation's infrastructure, and the systems accessed by users, from unauthorised access and potential damage. The main goal of cybersecurity is protecting an organisation's infrastructure and systems against any cyber-attacks and attackers. Most businesses depend on the internet to reach customers and provide services.Therefore, it is essential to define and implement measures to prevent cybercriminals from gaining access and stealing users' data and devices. Like individuals, organisations need to protect their IT assets from cyber-attacks taking place due to internal or external threats. Organisations need to convert unknown threats to known threats to identify and mitigate threats based on business risks. The relationship between defenders and attackers is asymmetric. The defenders need to prepare and be aware of all threats that may exploit their organisations' systems and infrastructure. On the other hand, the attackers need to exploit one vulnerability to gain access and cause damage to the organisation. This asymmetry gives the attacker a big head start comparing to the defender. The defenders need to collect information from all available sources, whether public or closed. For example, in 2020 [6] a threat actor was able to inject malicious code in the body of the update of the SolarWinds Orion [7], an IT system which helps to manage and monitor organisations' networks and infrastructure and is used by thousands of organisations. This system can give the attackers a complete view of those networks so they can steal sensitive information. Therefore, that attack enabled access to critical infrastructures and industrial control system organisations.

The attackers gained access to more than 18000 private and government organisations. They had the ability to take control of any affected installation because of malware installed in the previous version. The breach was discovered when a cybersecurity company called FireEye [8] which uses this software faced a breach. It had faced unauthorised access to its Red Team tools, a set of tools used by its security engineer team to exploit vulnerabilities in organisations' infrastructure. FireEye shared a list of countermeasures and rules in various standards to help the community detect the Red Team tools in their products and avoid any future attacks through these tools. Sharing information by affected organisations about the course of action and how to respond to this intrusion is essential to mitigate this risk for others. There are various potential sources of cyber threat data such as vendors, governments, private sources and open sources such as VirusTotal [9] and Cisco Talos Intelligence [10].

Cyber Threat Intelligence (CTI) has various definitions and meanings based on its goal and use. Henry Dalziel [11] indicated that cyber threat intelligence should have three main features: be relevant, actionable and valuable. Therefore, CTI should relate to the organisation's business and enable defenders to make meaningful and productive decisions. It should also be proactive instead of following a reactive tactic by providing awareness and insight about potential attacks.

In order to support sharing and analysing threat information, researchers and organisations are working to develop formats and standards for exchanging CTI information. The main standards can be listed as follows: Structured Threat Information Expression (STIX) [12] which is currently the most applied standard, Incident Object Description Exchange Format (IODEF) [13] and OpenIOC [14].

A reason for increasing the sharing of CTI information is the cost of data breaches, the number of attacks, and threat actors' motivations and capabilities [15]. Sharing helps organisations get better defence and increase threat detection accuracy [16]. For example, in [17], authors found that sharing URL lists related to malicious activity with hosting providers will minimise the possibility of using this URL to exploit systems. Besides, sharing this list contributes to blocking and stopping a malware attack quickly and effectively and identifying attack types.

CTI may contain sensitive and identifiable information about the victim's network infrastructure, existing vulnerabilities, credentials, business processes and financial information. As sharing information has become more common, privacy and confidentiality are considered to be a major concern and challenge. It is essential that evaluating the risk of sharing CTI is incorporated whenever sharing CTI is presented.

Privacy is a difficult-to-define concept across different communities. In the realm of laws and regulations, privacy may relate to personal information (e.g. an address or date of birth). In this thesis, we will use privacy to refer to any identifiable information in CTI datasets. There are various privacy-enhancing technologies (PETs) that can preserve privacy, confidentiality and mitigate potential threats against some adversaries. Different techniques have been proposed such as anonymisation techniques and models including k-anonymity [18][19] and differential privacy [20]. Chapter 2 provides a more in-depth exploration of the literature and privacy-enhancing technologies.

Sharing CTI datasets is a desirable action, but sharing without a qualified evaluation to quantify the risks of sharing CTI dataset would put the organisation at risk; for example, disclosing a vulnerability to the public will encourage attackers to exploit the systems especially when organisations require more time to patch their systems [21]. In the same context, organisations may well be justified in perceiving risks in sharing and disclosing cyber incident information, but they seem to be more reluctant to express such worries in clear and well-defined terms. Such concerns could also arise because of the risks of breaching regulations and laws in relation to privacy. With regulations such as the General Data Protection Regulation (GDPR) designed to protect citizens' data privacy, the managers of CTI datasets need clear guidance on how and when it is legal to share such information. Thus, it is paramount for CTI analysts and managers to understand

and delineate the legal risks and manage them by proposing a model to make the right decision of sharing.

However, evaluating the risk of sharing CTI datasets is challenging due to the nature of the CTI context, which is associated with the evolution of the threat landscape and new cyber attacks that are difficult to evaluate. Currently, a qualified evaluation remains unavailable. CTI managers face a tricky situation when deciding what to share, when, how and with whom. The scope of the challenge requires a coherent model that can assess the CTI dataset before sharing.

In this chapter, we introduce the research aims, research questions and summarise the major contributions and present the dissertation's outline.

1.1 Research Aims

The main aim of this research is to improve and stimulate cybersecurity information sharing while mitigating the potential adverse effects. We propose a model to quantify the risks and cover the legal aspects with adequate protection levels. This thesis proposes a framework to help an organisation to share cyber threat intelligence by proposing a risk assessment model validated by using several case studies. This framework will help organisations to make a decision about sharing this information. Organisations can be reluctant to share information without knowing the risks of sharing CTI dataset. The concrete objectives include the following.

1. Define and identify the risks of sharing cyber threat intelligence and determine the associated threats.
2. Define a risk assessment model to assess the associated threats of sharing cyber threat intelligence.
3. Define a model to evaluate and assess sharing cyber threat intelligence under the general data protection regulation (GDPR).

4. Validate the proposed models by applying them to case studies on cyber threat information sharing.

1.2 Research Questions

This research will explore the following questions:

1. What are the risks of sharing cyber threat information?
2. How to distinguish between sensitive and non-sensitive information in cyber threat information?
3. How to assess the risk of sharing CTI datasets caused by sharing with different entities in various situations?
4. How to evaluate the legal requirements for supporting decision making when sharing CTI?
5. How can the risks of sharing CTI datasets be mitigated?

1.3 Contribution

This thesis presents five main contributions by answering the research questions in Section 1.2. The contributions of this thesis are the following:

1. Detailed analysis and understanding of which information in cyber incident reports requires protection against specific threats with assessed severity
2. A quantitative risk model to assess the risk of sharing CTI datasets enabled by sharing with different entities in various situations
3. A comprehensive evaluation of the risks of sharing cyber incident report and the quantitative risk model using three practical use cases

4. A set of guidelines for disciplined use of the STIX incident model in order to reduce information security risk

5. A model for evaluating the legal requirements for supporting decision making when sharing CTI, which also includes advice on the required protection level

1.4 Structure of the thesis

To address the research questions described in Section 1.2, this thesis is organised as follows.

- Chapter 2 reviews the literature on privacy-preserving techniques, including anonymisation, encryption and differential privacy. It introduces Cyber Threat Intelligence (CTI), its definition, lifecycle, subdomains, standards and challenges. Furthermore, it explores CTI standards focusing on the STIX standard, threat types and the legal requirements on cyber information sharing. Besides, it examines risk assessment approaches and discusses related work in the context of risks of sharing CTI information and the legal aspects.

- Chapter 3 answers the first research question by performing an exploratory study on the threats of sharing cybersecurity incident information and provides a detailed analysis and understanding of which information in cybersecurity incident reports requires protection. First, it defines the sensitive information and the identification categories of the incident model attributes and defines a severity analysis of threats based on a threat taxonomy. Second, it calculates information disclosure threat of the STIX incident model. Finally, the chapter extends our method to other standards of cyber threat information.

- Chapter 4 answers the second and third research questions by proposing a quantitative risk model to assess the risk of sharing CTI datasets among different entities in various situations. It defines the associated risk model by elaborating dataset analysis, threat analysis and total associated risk. Finally, Chapter 4 presents the evaluation of our analysis in and the risk model evaluation through three use cases that help determine the risk level of sharing a CTI dataset and consequently, the mitigation techniques to enable responsible sharing. All use cases have been validated using professional and academic experts' opinions.

- Chapter 5 answers the fourth and fifth research questions by defining the impact that GDPR legal aspects may have on the sharing of CTI. In addition, it defines a flow diagram related to cybersecurity information sharing and adequate protection levels for sharing CTI to ensure compliance with the GDPR. It also presents a model for evaluating the legal requirements for supporting decision making when sharing CTI, which also includes advice on the required protection level. Finally, we evaluate the model through two use cases of sharing CTI datasets between different entities and discuss the results.

- Chapter 6 concludes the thesis and discusses outstanding research issues for future research.

1.5 List of publications

The following publications are part of this thesis, and the first author is the principal author who undertook the majority of the underlying research as well as the production of the papers.
Chapter 3 is based on the work published in the following paper.

- Albakri, A., Boiten, E. and De Lemos, R., 2018, August. Risks of sharing cyber incident information. In Proceedings of the 13th International Conference on Availability, Reliability and Security (pp. 1-10).

Chapter 3 has been extended by presenting the full table analysis and expanding to other sharing cyber threat information standards.
Chapter 4 is based on the work published in the following paper.

- Albakri A., Boiten E., Smith R.,2020. Risk Assessment of Sharing Cyber Threat Intelligence. In: Boureanu I. et al. (eds) Computer Security. ESORICS 2020. Lecture Notes in Computer Science, vol 12580. Springer, Cham. https://doi.org/10.1007/978-3-030-66504-3_6

The Chapter has been extended by presenting additional experiments and evaluating an open-source STIX dataset and an additional use case study.
Chapter 5 is based on the work published in the following paper.

- Albakri, A., Boiten, E. and De Lemos, R., 2019, June. Sharing Cyber Threat Intelligence Under the General Data Protection Regulation. In Annual Privacy Forum (pp. 28-41), vol 11498. Springer, Cham. https://doi.org/10.1007/978-3-030-21752-5_3

Further investigations have extended the chapter to gain better understanding and insight into the legal dimension.

Chapter 2

Background and Related Work

This chapter serves as a background for the remainder of the thesis. We first look at the concept of privacy, then we study privacy preserving techniques. After that, we delve into cyber threat intelligence and its standards, challenges and legal requirements. Finally, we discuss a research area related to risks of sharing cyber threat information, risk assessments and sharing cyber threat intelligence under laws and regulations.

2.1 Introduction to privacy

Based on the context and the community, there are several meanings and understandings of privacy. Therefore, there is not a specific meaning accepted as the only definition. The general definition based on [22] is “someone's right to keep their personal matters and relationships secret”. The United Nations has defined privacy as one of the fundamental human rights (Art.12) “no one shall be subjected to arbitrary interference with his privacy, family, home or correspondence” (United Nations, 1948). On the other hand, in [23] Brandeis and Warren are the first authors who highlight privacy as a right and define it as “the right to be let alone”. Also, [24] propose a more specific definition of the privacy which is “the right of the individual to decide what information about himself should be communicated to others and under what circumstances”. In this definition, they give the user the right to control which information they want to share, with whom and how. Moreover, in [25] Solove defines privacy as “a concept in disarray. Nobody can articulate what it means. As one commentator has observed, privacy suffers from an embarrassment of meanings.”. Solove has published a taxonomy of privacy harm aiming to classify the harms that may occur from privacy violation. Solove aimed at providing further understanding of privacy violations in various contexts. The main harmful activities Solove included are the following:

Information collection: for example, surveillance which includes monitoring all user’s activities, and interrogation consists of various forms of questioning or probing for information. This can be seen when starting with network traffic surveillance and then analysing network traffic. The analysis might lead to determine what type of data is being exchanged, and the kind of network protocols that are in use [26].

Information processing: In this group, there are activities related to how data is stored, operated and used. Aggregation includes a grouping of different pieces of data about an individual. Hence, identification is linking information to a specific person. Insecurity includes the failure to protect the data against a data breach or unauthorised access. A secondary use is when using the data for a different purpose without the data subject’s consent. Finally, exclusion entails preventing the data subject from knowing about the data others store about them.

Information dissemination: in this group, the main activities include breach of confidentiality which is failing to keep the information confidential. Disclosure includes revealing information that might make an impact on someone. Exposure includes exposing specific physical and emotional characteristics about an individual to others. Increased accessibility involves expanding the accessibility of information. Blackmail is the activity of threatening someone to reveal a piece of personal information. Appropriation includes perpetrating identity theft to serve the aims of another. Finally, Distortion entails the spreading of false information about individuals.

Invasion: this group of activities include intrusion which is an action that could interrupt someone's life activities and Decisional interference, which includes the authority's infiltration into the data subject's decisions about their personal life.

Also, Nissenbaum [27] introduced a framework and described "contextual integrity" as a new standard of privacy. Privacy is described as "Privacy as appropriate information flow". According to this definition, the appropriate flow means that the flows conform to the rules that meet the expectations of different social domains. Also, the appropriate flow requires that information collecting and sharing be appropriate to that domain and follow its governing rules of distribution. Based on this privacy definition, the information flow considers three parameters. The first parameter is "Actors" which is "subjects, senders and receivers" of the data. For example, the value of this parameter could be a hospital, insurance company, central authority, a teacher or a friend. The second parameter is "Information Types" which represents the ontology of attributes that are related to the context, for example, Social Security number (SSN), gender, salary or facial images. Finally, "Transmission" which refers to how the data is shared between actors, for example, consent, hack or sell. In respect to these definitions and understandings of privacy, it would seem that Nissenbaum's definition would align satisfactorily with this thesis. The researcher looked at the data flow processes and classified them into different classes. Also, they looked at how the information transmitted and covered various aspects, such as sharing information with a third party. Sharing information explicitly contains privacy risk. Consequently, sharing needs to be secure and needs to maintain privacy requirements.

2.2 Privacy and analysis approaches

Essentially, a provider has two primary methods of sharing access to data with others in a privacy-preserving way: they can share a modified copy, or they can give modified access to the original.

The conventional approach of sending a modified copy is by applying anonymisation methods or de-identification, so the provider does not need to trust the analyst. The definition of data anonymisation is “process by which personal data is irreversibly altered in such a way that a data subject can no longer be identified directly or indirectly, either by the data controller alone or in collaboration with any other party.” [28]. In other words, anonymisation implies removing any directly identifying personal data such as name and national insurance number from the individual’s record. However the negative aspect of this approach is that the analyst does not have access to the rich, full data, so they may not find the conclusion and the knowledge they want from the data. Moreover, this approach does not provide a satisfactory result to protect an individual’s personal data.

Aggregate statistics as a privacy solution is not successful because of differencing attacks [29]. For example, a query against large dataset could identify specific individuals. Let us assume we know that Bob is in a particular health dataset. Then the following two questions could confirm whether Bob has heart disease “How many people in the dataset have heart disease” and “How many people whose name is not Bob in the dataset have heart disease”. Therefore, to avoid difference attacks, auditing and observing the sequence of user queries and responses could be the answer. However, the challenges of query auditing mainly come from the idea that it is possible not to answer the query. The query is a sign that the individual exists in the dataset. Besides, the query auditing algorithms cannot be accurate in deciding if a sequence of queries establishes a differencing attack [29]. Furthermore, other problems come through different reconstruction attacks [30] against a dataset where each person has a “secret bit” to be protected. Revealing “ordinary” information might be problematic if an individual is following specific behaviour over time, such as buying bread and then stopped this action. Therefore, an analyst might conclude that specific individual, in the dataset, is trying to lose weight, and that could be harmful to that individual.

On the other hand, giving the analyst supervised access to the dataset will produce better research results. At the same time, the analyst must convince the provider that the analysis is respectful of individual privacy. The differential privacy [29] approach also works in this context, potentially distorting the data to provide privacy at a minimal expense to accuracy. Differential privacy offers a way to protect all data subjects, even the outlier individuals in the dataset whose privacy could be at risk due to various statistical attack types. This model allows the data analyst to run queries adaptively, deciding the level of accuracy of the answer. All these approaches will be discussed in more detail in the following sections.

2.3 Privacy preserving techniques

Many organisations and private companies want to share an anonymised version of their data with the public, for example, as a part of an “open data” agenda. Some attempts of using privacy-preserving techniques such as de-identification have notoriously gone wrong. For example, on August 9th, 2006, the technology section of the New York Times contained a news item entitled “A Face Is Exposed for AOL Searcher No. 4417749.” [31]. AOL Research had published a dataset containing internet search records of 650,000 customers for the period of three months, adding up to 20 million search queries. This sharing was intended for academic research benefits. They replaced most identifying data, but sometimes if users look at their search history, they may find their data such as names, addresses or any personal information. After a short period, the New York Times published one identified individual's information from that data set. For example, over three months period, the user No.4417749 conducted a series of queries containing information such as “numb fingers” to “60 single men” to “dog that urinates on everything.” After following the queries of this user, it became easier to identify this person especially when geographical information has been stated such as “landscapers in Lilburn, Ga,” and “homes sold in shadow lake subdivision Gwinnett county Georgia.”. Ultimately, the researchers were able to identify the user.

In another example, as part of the Netflix Prize contest, Netflix -the world's biggest streaming media service- publicly published a dataset consisting of movie ratings of 500,000 Netflix customers. The prize was $1 million for developers who can enhance the accuracy of the company's current movie recommendations system based on personal viewing and movie rating history [32]. The dataset was planned to be anonymous, and all personally identifying information had been extracted. In [33] they were able to identify users in the Netflix database by using the Internet Movie Database (IMDb) from imdb.com as an external data set. They proposed an algorithm that can be used against any dataset containing anonymous individual records such as transaction and preferences. Therefore, removing identifying information is not enough for anonymity and does not provide enough privacy guarantee. Targeted re-identification could occur after a normal conversation with a colleague at work about movies they watched, and their rating of these movies. This action could put their privacy at risk. Researchers found how much the attacker needs to know about a Netflix customer to identify their record if it exists in the dataset.

2.3.1 Anonymisation methods

Existing models of anonymisation divide data into different types of attributes. These attributes affect data anonymisation/de-identification techniques. For a disclosure to be harmful, it needs to imply 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. The attributes' types are:

Identifier attributes include information used to directly identify an individual such as name, passport number, and national insurance 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 and salary.

There are three basic criteria for checking the quality of models used to remove personal information from an individual's record in a way that decreases the possibility of disclosure of the identity of individuals. These models are k-anonymity [18], l-diversity [34] and t-closeness [35].

k-anonymity: One of the most well-known anonymisation mechanisms applied to data is k-anonymity [18][19] . For any dataset that satisfies k-anonymity, each record cannot be distinguished from at least k-1 records of information that appear in the dataset. The main techniques for achieving k-anonymity for some value of k are:

Suppression: In this technique, we replace certain value of the attributes by an asterisk '*', and that could be applied to part or all the value of the attribute. Table 2 replaced part of the value in the 'Age' column by '*'. Many algorithms have been designed for suppression, such as the Truncation algorithm [36] which might be used to anonymise the IP and MAC addresses by deleting the first bits and replacing the removed bits by zeros. We may use different algorithms based on the type of data stored. For example, the Black marker algorithm [37] removes and replaces any field with a fixed value, but like most anonymisation methods, that will reduce the usefulness of the dataset. On the other hand, the Enumeration algorithm [36] starts by sorting the data then selecting a value higher than the first value and adds this value to all fields. This algorithm can be applied only to numeric fields [36]. These algorithms come close to perturbation which we will discuss in more detail in Section 2.3.2.

Generalizations: In this technique, we group values together. The idea is based on converting the values of the attributes of a specific domain by a more general value. For example, the value '31' of attribute age would be replaced by [25-35]. Some algorithms are designed for time, such as the random time shift algorithm [36] that adds a random offset to the timestamp attribute. For example [35], Table 1 shows the health information for the patients in the hospital. The Table contains quasi identifier attributes (Zip code and Age), and the sensitive attribute is the disease. Table 2 shows a 3-anonymity version derived from Table 1. The "*" refers to a suppressed value, for example, "age =2*" represents the age between 20 and 29.

	ZIP Code	**Age**	**Disease**
1	36677	28	Heart Disease
2	36602	21	Heart Disease
3	36678	26	Heart Disease
4	36905	42	Flu
5	36909	51	Heart Disease
6	36906	46	Cancer
7	36605	30	Heart Disease
8	36673	35	Cancer
9	36607	31	Cancer

Table 1: Original Patients Table

ZIP Code	**Age**	**Disease**
366**	2*	Heart Disease
366**	2*	Heart Disease
366**	2*	Heart Disease
3690*	$\geqslant 40$	Flu
3690*	$\geqslant 40$	Heart Disease
3690*	$\geqslant 40$	Cancer
366**	3*	Heart Disease
366**	3*	Cancer
366**	3*	Cancer

Table 2: A 3-Anonymous Version, QI = {Zip Code , Age }

k-anonymity cannot provide full protection from sensitive attribute disclosure. There are attacks on k-anonymity, such as Background Knowledge and Homogeneity Attack when there is a lack of diversity of insensitive attributes [33]. For example, in Table 2 Alice and Bob are neighbours. One day Bob went to the hospital and Alice wanted to discover what Bob's disease was and she had access to the 3-anonymous Table. She already knew that he is living in the zip code 36677 and is 28 years old and since all the patients have the same disease, which is Heart Disease, she found out that Bob has Heart Disease as well. Accordingly, to prevent Background knowledge and Homogeneity attack, an extended technique has been proposed: l-diversity.

l-diversity [34] is an extension of k-anonymity to protect the published data against Background Knowledge and Homogeneity Attack. l-diversity ensures that not only all users are k-anonymous, but also each group of users shares a variety of sensitive information. The variations of sensitive attributes ensure that all sensitive attributes are adequately distributed to avoid attribute disclosure.

To achieve l-diversity one may need to insert dummy data to increase the variation of sensitive information, hence, extracting useful information may be a big challenge. Also, l-diversity is subject to various types of attack that could cause an attribute disclosure [35]. The first attack is Skewness attack. It is an attack when the overall distribution is skewed and does not consider the overall distribution of sensitive values. Let us assume that we have a single sensitive attribute with two values, and the level of sensitivity is different between those values. Then the classes present different levels of privacy risk.

The second attack is similarity attack when the values of the sensitive attribute in an equivalence class are different but similar in a semantic way. In this case, an attacker can infer important information about individuals. For example [35], Table 4 shows the 3-diverse version of the Table 3. Comparing with the 3-anonymous Table Alice cannot know from the database the association of Bob's record and his sensitive attribute value. But if intruder knows that Bob's record related to the first group, then we can infer that Bob has stomach-related disease and his salary to some extent low.

As a result, it has been argued that it is challenging to achieve l-diversity, and even reaching l-diversity may be inadequate to avoid attribute disclosure. Besides, it becomes more challenging to achieve l-diversity when there are multiple sensitive fields [38].

	ZIP Code	**Age**	**Salary**	**Disease**
1	37677	29	3K	Gastric ulcer
2	37602	22	4K	Gastritis
3	37678	27	5K	Stomach cancer
4	37905	45	6K	Gastritis
5	37909	54	11K	Flu
6	37906	48	8K	Bronchitis
7	37605	33	7K	Bronchitis
8	37673	32	9K	Pneumonia
9	37607	31	10K	Stomach cancer

Table 3: Original Salary/Disease

ZIP Code	**Age**	**Salary**	**Disease**
376**	2*	3K	Gastric ulcer
376**	2*	4K	Gastritis
376**	2*	5K	Stomach Cancer
3790*	$\geqslant$ 40	6K	Gastritis
3790*	$\geqslant$ 40	11K	Flu
3790*	$\geqslant$ 40	8K	Bronchitis
376**	3*	7K	Bronchitis
376**	3*	9K	Pneumonia
376**	3*	10K	Stomach Cancer

Table 4: 3-diverse version of Table 3

t-closeness [35] t-closeness is an-other extension of l-diversity that decreases the granularity and makes the distribution of the sensitive attribute in any equivalence class close to the distribution of the entire attribute. In this approach, researchers measured privacy based on the information gain of an observer. They proposed the information gain as the difference between the observer expectation of the sensitive attribute value of an individual before releasing the data and the posterior expectation after releasing the data and seeing the value.

The novelty of this approach is mainly separating the information gain into two parts: the first is about the full attribute values in the released data, and the second is about specific individuals. The definition of t-closeness principle is "An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the whole table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness" [35]. For example [35], Table 6 shows 0.167-closeness with regard to Salary and 0.278-closeness with regard to Disease anonymisation derived from Table 5.

	ZIP Code	**Age**	**Salary**	**Disease**
1	37677	29	3K	Gastric ulcer
2	37602	22	4K	Gastritis
3	37678	27	5K	Cancer
4	37905	45	6K	Gastritis
5	37909	54	11K	Flu
6	37906	48	8K	Bronchitis
7	37605	33	7K	Bronchitis
8	37673	32	9K	Pneumonia
9	37607	31	10K	Cancer

Table 5: Original Salary/Disease

ZIP Code	**Age**	**Salary**	**Disease**
3767*	$\leqslant 40$	3K	Gastric ulcer
3767*	$\leqslant 40$	5K	Cancer
3767*	$\leqslant 40$	9K	Pneumonia
3790*	$\geqslant 40$	6K	Gastritis
3790*	$\geqslant 40$	11K	Flu
3790*	$\geqslant 40$	8K	Bronchitis
3760*	$\leqslant 40$	4K	Gastritis
3760*	$\leqslant 40$	7K	Bronchitis
3760*	$\leqslant 40$	10K	Cancer

Table 6: Table that has 0.167-closeness with regard to Salary and 0.278-closeness with regard to Disease

In this method, the limitation that requires the distribution of a sensitive attribute in any equivalence class is close to the distribution of a sensitive attribute in the overall table, which constitutes a challenge with multiple sensitive attributes. Moreover, the relationship between the value t and information gain for measuring the privacy is vague. From the above three mentioned techniques, we can see that the anonymisation techniques preserve the consistency of the database, and all the operations are at the record level. All the previous techniques try to reduce information loss to make the released data more useful. To achieve this, many algorithms focus on the information loss of the released dataset. Many information loss metrics have been proposed, such as The Classification Metric (CM)[39], The Discernibility Metric (DM) [40] and the Generalized Loss Metric [39]. Some of these measures are suitable for specific data mining algorithms and cannot be used for general applications. In addition, [41] proposed a framework to identify the utility of attributes in a data set. Choosing the correct anonymisation methods depends on the dataset and the types of attributes such as quasi-identifier and sensitive attributes.

In [42] researchers proposed a cyber threat intelligence framework prototype. The main component in that framework is an anonymisation tool. The purpose of this tool is to anonymise attributes such as IP addresses, NI numbers and E-mail addresses. They used specific regular expressions to extract identifiable information from the dataset. They defined a set of anonymisation rules which can be activated based on different anonymity level depending on the TLP protocol [43] where they can be modified according to the organisation's requirements.

2.3.2 Perturbation

Perturbation techniques attempt to preserve privacy of the data by applying randomized modification to the sensitive attributes to hide them, usually utilising additive noise [44]. One of the simple ideas is to replace the original values with perturbed values that maintain the correctness of the important overall properties of the dataset.

Applying perturbation techniques will work best for numerical data mainly when most of the perturbation methods focus on operations such as swapping, updating and deleting values of the original data [45]. For example, swap between data records maintains the statistical quantities such as aggregate counts, average and distribution of the data.

Table 8 is an example of perturbation operation Micro-aggregation [46] where it replaced the original data by the average computed on a small group of records, The records belonging to the same group will be represented in the disseminated information by the same value.

In the perturbation technique, it is impossible to restore the original data. Various methods have been proposed to measure the utility of the shared data by evaluating the distance between the released data and the original data [46]. The limitation of these methods is that it is hard to apply them automatically with

hid	**Income**
8393	1360
3236	4243
7188	11163
9503	18145
9204	25149
2866	26310
5386	32538
8787	32600
6376	35300
7781	36099
3672	37228
5089	38001

Table 7: Original Table

hid	**Income**
8393	**5588.49**
3236	**5588.49**
7188	**5588.49**
9503	23291.45
9204	23291.45
2866	23291.45
5386	**33479.17**
8787	**33479.17**
6376	**33479.17**
7781	37109.44
3672	37109.44
5089	37109.44

Table 8: Perturbation operation result

non-numerical data.

2.3.3 Differential privacy (DP)

Differential privacy (DP) [20] [47] is a mathematical definition of privacy and an approach which represents preserving privacy. For any algorithm considering differential privacy, the output of this algorithm will be independent of whether any individual's personal information joins or leaves the statistical dataset. Differential privacy will give the ability to extract statistical data from a dataset containing personal information without disclosing this information. Moreover, differential privacy solves the problem of database linkage attacks. The mathematical definition of differential privacy is [20] :

$\mathcal{M}$:$\mathcal{X}^n\times$ $\mathcal{Q}\rightarrow y$ is ε-differentially private if for every pair of neighbouring datasets $x \sim x^{'} \in \mathcal{X}^n$ (i.e., x *and* $x^{'}$ differ in one row), and every query $q \in \mathcal{Q}$, we have:

$$\forall\ T\ \subseteq Y, \Pr\left[\mathcal{M}\left(x,q\right) \in T\right] \leq e^{\varepsilon}\ .\Pr\left[\mathcal{M}\left(x^{'},q\right) \in T\right] \tag{1}$$

The value of ε is small; for example, ε= 0.1, a smaller ε provides better privacy.

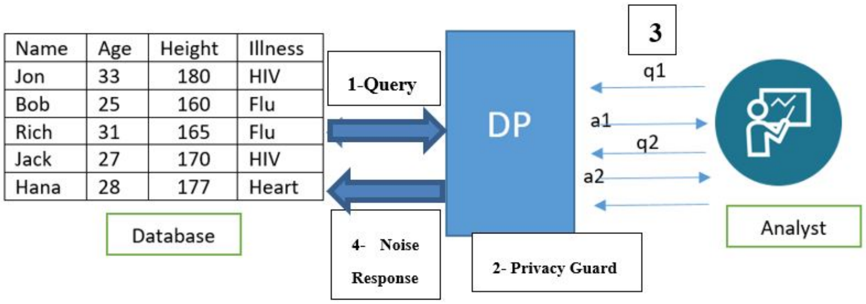
In DP, we have a 'privacy budget' measure which is given based on the value of ε. The value of ε will be increased after every query and when the budget value is exceeded, the user will not have access to run any queries on the dataset. Figure 1 illustrates how the steps for DP can be applied to big data [1]. In the process, there is no direct access to the database that contains the original data, but there is an intermediate layer called DP privacy guard between the analyst and the database to preserve the privacy. The steps are:

1. The analyst is able to send a query to the database through the DP guard.
2. The guard checks the query the analyst wants to ask of the database and measures the privacy consequences of this query combined with the queries

that have come before it. This evaluation relies on the sequence of the queries the analyst asks without considering the real data in the database.

3. The guard sends the request to the database and receives the response without noise.

4. The guard adds noise to the query based on the privacy risk and sends the new result to the analyst.



Name	Age	Height	Illness
Jon	33	180	HIV
Bob	25	160	Flu
Rich	31	165	Flu
Jack	27	170	HIV
Hana	28	177	Heart

Figure 1: Differential privacy mechanism [1]

To be able to apply the queries against the full database without the need for the providers' trust or understanding the analysis, a new programming language for differential privacy, Privacy Integrated Queries (PINQ) [48], has been proposed. The analysts write PINQ queries analogous to writing to the database and get the aggregate result after the noise has been added to it.

2.3.4 Encryption–based methods

The goal of this technique is to enable computation when sharing information while preserving the privacy of participating parties. A number of cryptographic approaches have been proposed for ensuring the confidentiality and privacy of the shared data, but existing techniques are greedy in resources. In this section, we will describe the existing cryptographic approaches designed for ensuring privacy.

The global approaches are based on homomorphic encryption and multiparty computation that enforce privacy by design. The measure of achieving confidentiality and privacy depends on adversary models. Cryptographic approaches can deal with different adversarial models; there are two main types of adversary models: honest-but-curious (HBC) and malicious adversary.

Honest-But-Curious (HBC) Adversary: In this model, all parties act exactly as the honest party and follow the actions in the protocol. However, the adversaries try to know more about the private information of other parties.

Malicious adversary: In this model, the adversaries may deviate arbitrarily from the specified actions in the protocol such as sending distorted messages or actively colluding with other malicious parties to violate the privacy or integrity of the others players' private data.

Homomorphic encryption [49] is a type of encryption that allows computation on encrypted data and generates an encrypted result that matches the result of operations on original data. This will preserve the confidentiality and the privacy of the data during the operations. There are many applications which can employ Homomorphic encryption schemes such as cloud computing [50].

As a definition [51], consider a cryptosystem C has an encryption function, plaintext X_n, and some operation $\triangle$.

C is considered additively homomorphic if:

$$\exists\Delta : \ \varepsilon\left(x_1\right)\Delta\varepsilon\left(x_2\right) = \ \varepsilon(x_1 + x_2) \tag{2}$$

C is considered multiplicatively homomorphic if:

$$\exists\Delta : \ \varepsilon\left(x_1\right)\Delta\varepsilon\left(x_2\right) = \ \varepsilon(x_1 x_2) \tag{3}$$

Therefore, we can apply this definition to any other operation. There are two forms of homomorphic encryption: Fully homomorphic encryption and partially homomorphic encryption. Fully homomorphic encryption (HE)[49]: fully HE supports arbitrary computations on encrypted data. No information will be revealed because both the input and output are encrypted. There are a variety of fully HE practical uses in applications such as, consumer privacy in advertising when collecting information about the user for advertising purposes which trigger privacy concerns for example, to solve geographic targeting [52] and avoid monitoring the user's habits. When the provider collects geographic information, homomorphic encryption could send the advertisements as they come from a third party and without monitoring the advertising provider. Also, [53] presented a recommendation system where the user receives a computed recommendation using HE without revealing the user preferences to the system. Health applications can use Fully HE techniques [52] by allowing health care providers to encrypt the patient's medical record and then uploading it to the cloud storage system. Only the user can decrypt the data, and at the same time computing can be done on users' encrypted data that contains blood pressure, heart rate, and other health measures to keep tracking the patient's health. There are limitations to fully HE when supporting multiple users with multiple encryption keys and huge computations. In addition, it is currently inefficient to execute complex algorithms in a homomorphic way.

Partially Homomorphic Encryption [53] [49] is executing one operation such as addition or multiplication but not both. One application of partially homomorphic encryption is RSA that exhibits multiplicative homomorphism. As a summary, we can preserve privacy and confidentiality when applying homomorphic encryption, and it can be used in many practical applications, but still, the main challenge of HE is the complexity of the systems and the performance.

Secret sharing by [54] is a set of encryption based methods that enforce privacy and availability by design. Secret sharing distributes a secret among n participants where each participant will receive an unintelligible share of the data. The

secret will be reconstructed only after combining enough parts together through a trusted client. Extension schemes of the secret sharing have been developed, such as multi-secret sharing [55] and verifiable multi-secret sharing [56] to improve the existing protocols and provide solutions against attacks.
Secure multi-party computation (MPC) was introduced by [57]. In MPC each party holds some private data, the parties will perform multi-party computation on this private data and only the receiver can reconstruct the output. One of the methodologies for secure MPC is secret sharing. There are many practical applications of MPC [58] such as collecting and analysing financial data for consortium of information and communication technology companies [59].
In [60] researchers provide a comprehensive analysis of the mechanisms and challenges of privacy preservation in big data focusing on the infrastructure and all big data life cycle stages such as generating, storing, and processing. In the data generation phase, the main techniques restrict the access or distort the data. Some techniques and tools could be used to distort the data such as MaskMe [61] for hiding online identity.

We need to ensure that the data is secure against any disclosure threat in the data storage phase in distributed environments. The techniques used in this phase are encryption techniques and the current approaches are:

- Attribute Based Encryption (ABE) [62] [63] which is a type of encryption that ensures the privacy of big data storage systems by providing fine-grained and flexible access control. To decrypt the data in this type, we need a set of attributes and private key matching the information that is associated with the encrypted data.

- Identity Based Encryption (IBE) [64] which is an alternative scheme to public key cryptosystems where the public key could be any string, such as email address or IP address.

- Homomorphic Encryption (HE) in which one can perform operations on the encrypted data.

- Storage Path Encryption [65] In this scheme, the user will use a trapdoor function to secure storage of data. In this approach, instead of encrypting the big data itself we just encrypt the storage path.

Organisations might share private sets for union, intersection, and difference operations. This sharing brings privacy risks such as disclosing organisations' set instead of getting the result of multiset operations. In [66], researchers proposed a framework for privacy-preserving operations such as union, intersection and element reduction. One practical example of set interaction problem is 'do-not-fly' list that requires private intersection between the government's list and the airline's passenger lists.

2.3.5 Sticky policy and Privacy patterns

One definition of sticky policy is "machine readable policies that can stick to data to define allowed usage and obligations as it travels across multiple parties enabling users to improve control over their personal information" [67]. In sticky policy, an obligation management system in service providers manages information lifecycle management depending on personal preferences and organisational guidelines. In many situations, a company needs to reveal personal or maybe sensitive information in order to get a specific service. However, achieving the goal of sharing needs a mechanism to ensure addressing all policies. The main characteristics of the sticky policy are: define the purpose of sharing, such as research, use the data within a specific technical environment, define what they can and cannot do with the data, define the retention policy and finally, define the list of trusted authorities that can provide assurance and accountability in the procedure of giving access to the protected data. Therefore, with these features, the company will be able to define how their data should be processed, stored and shared by defining their conditions explicitly.

The main advantages of sticky policy according to [68] are as follows: the data owner can set and manage their preferences on their data before sharing them with others because the policy transfers with the data, protecting the entire data life cycle. The management of the policies and access control would be easier since the third party will be responsible for supervising and managing policy enforcement systems. Besides these advantages, there are shortcomings. It is difficult to provide an adequate set of policies when data is coming from different domains, with different formats and semantics. Consequently, it is challenging to develop a standard. The computational cost of processing and transmitting the policies among the data is high. The main shortcoming is that sender needs to trust recipients to respect the sticky policy. In addition, to address one of the previous challenges, many models have been proposed for defining sticky policies associated with data sharing agreements [69] [70]. These models introduced a novel way to represent sticky policy generically and structurally.

Privacy patterns are another way of ensuring privacy and providing practical guidance for software engineering. To represent privacy concerns among different parties we need to provide privacy patterns that might help to standardize language for privacy preserving technologies, identify the standard solution to privacy issues, and help designers to pinpoint and deal with privacy concerns [71] [72]. As Alexander in 1977 wrote about patterns in general [73] “Each pattern describes a problem which occurs over and over again in our environment, and then describes the core of the solution of the problem, in such way that you can use this solution a million times over, without ever doing it the same way twice”. There are different types of patterns related to different issues and privacy problems introduced such as onion routing for anonymous communication, obligation management pattern and aggregation gateway pattern. The obligation management pattern permits obligations in relation to data sharing, storing and processing to be transferred and managed when sharing data between various entities. The “Aggregation Gateway” pattern allows encryption, aggregation and decryption at different places to prevent revealing personal information when computing specific measurements of a service. This model tries to solve the problem of protecting the information of the data subject. It is helpful, especially when the system needs computation and aggregation of measurements of a service attribute linked to a consumer for example, when an electric power grid installs smart meters to provide measurements of the current power consumption of each consumer. The provider will be able to achieve optimal energy consumption and update the power distribution dynamically, based on user necessity, without violating user privacy.

2.4 Cyber threat intelligence

Cyber threat intelligence is necessary for governments and organisations to protect systems and critical infrastructures, and to ensure the security of national services such as public health and defence. In addition, it helps to reduce the uncertainty related to cybersecurity investment [74]. The increasing number of cyber-attacks and the changing landscape of cyber threats have made the need to collect and analyse cyber threat information critical for defending against security incidents and data breaches [75]. The number of information leakages is increasing. In some cases, data leakage prevention tools leak information [76]. Analysing this leaked information allows the detection of those data breaches and finding correlations to discover potential leaks. It provides insights into the behaviour of similar attacks and helps security analytics operations to resolve incidents more quickly. In [77], the authors proposed a framework to detect and process potential leaks of critical information to reduce manual analysis conducted by CSIRT operators and analysts. This framework analyses unstructured and structured feeds from various sources. It extracts identifiable and sensitive information such as credentials and cards information. It classifies information based on defined taxonomy and allows analysts to search and explore data. Cyber threat intelligence is not limited to incident reports or low-level technical attributes. Other useful types of data to be exchanged include threats, vulnerabilities, mitigations, situational awareness, best practices and strategic analysis. For example, on November 24, 2014, Sony Pictures Entertainment was hacked, and sensitive information was leaked including business and personal information [78]. In order to stop similar upcoming attacks, corporate and governmental collaboration started to share threat information and courses of action [79] [80]. The sharing was mainly based on indicators of compromise (IoCs) and low-level technical attributes. After that, in 2015 and 2016, a cyberattack using the SWIFT banking network was reported with a financial loss valued at millions of dollars [81]. The correlation between those two attacks was not based on the IoCs, but it was based on the tactics, techniques and procedures used to achieve the attack's goal. Both attacks were assigned to the same threat actors, called the Lazarus Group [82] [83]. This shows that sharing cyber threat information between organisations is essential in order to help analysts to discover the intrusions and who may be behind the attacks. Also, it helps organisations to manage the risks and improve against future attacks. Security software vendors provide regular overviews on their websites or in "white papers" about how to exchange and share security events, intelligence and technical details on latest cyber-attacks and advanced persistent threat groups. Accordingly, cyber threat intelligence has become an essential component of organisations' security structure.

There are many definitions of cyber threat intelligence. From 2013, an illustrative definition of cyber threat intelligence by McMillan is "Threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard" [84]. [85] proposed another definition of cyber threat intelligence derived from a definition of intelligence: "Intelligence is the collecting and processing of that information about threats and their agents which is needed by an organization for its policy and for security, the conduct of non-attributable activities outside the organisation's boundaries to facilitate the implementation of policy, and the protection of both process and product, as well as persons and organizations concerned with these, against unauthorized disclosure". Thus, cyber threat intelligence is defined as "Cyber Threat Intelligence is nothing more than the application of intelligence principles and tradecraft to information security. Its outcome is nothing different from traditional intelligence: to inform and empower decision-making at all levels with knowledge of threats." Another definition was introduced by Lee [86] as "The process and product resulting from the interpretation of raw data into information that meets a requirement as it relates to the adversaries that have the intent, opportunity and capability to do harm."

In summary, all definitions endeavour to delineate the purpose of threat intelligence in different aspects. For the rest of this thesis, we will rely on McMillan's definition as it is the most comprehensive.

2.4.1 Cyber threat intelligence lifecycle

The intelligence process life cycle consists of the following six phases [87]: planning and direction; collection; processing and exploitation; analysis and production; dissemination and integration; and evaluation and feed-back.

Planning and direction: This phase includes various tasks such as the identification and prioritisation of intelligence requirements. It determines the goal for collecting this intelligence, whether strategic, technical, tactical or operational and defines a collection plan for the intelligence. Another task is designing an intelligence team by determining the personnel and equipment. It includes sending requests for data gathering from internal and external sources and to develop requirements for other phases.

Collection: This phase focuses on the activities of collecting the intelligence defined in phase one. The source of this information could be human or technical. Based on the type of information, it could be collected directly from public sources such as public blacklists [88] or through intermediary channels which act as trusted entities such as national CERTs. This intelligence includes collecting information from log data, network traffic and monitoring system. During this phase, the effectiveness of the collecting plan will be evaluated. Then, the data will be transferred for processing in the next phase.

Processing and exploitation: the data collected from the previous stage is not in a standard format, and it needs to be transformed to be used by analysts and decision-makers. In this phase, experts use specific tools and methods such as decryption, filtering, aggregation and language processing to convert this data into a standard structure format. A clear collection plan and requirements of the user and nature of the processed data make the processing effective.

Analysis and production: this phase aims to produce intelligence from the information collected in the previous stage and convert it into accurate and useful information. The analysis includes proofs, outcomes, and forecasts, which help the evaluation process and prediction of future attacks. The results should provide the right information at the right time and need to be actionable. There are various types of analysis to produce intelligence at this point, such as quantitative and qualitative, machine-learning methods and statistical analysis. This phase tries to identify potential threats to the organisation and provide supports in prioritising controls to mitigate and evaluate the identified threats.

Dissemination and integration: During dissemination and integration, the

result of the previous phase will be delivered to and consumed by the intelligence consumer automatically or manually. These results include context on indicators of compromise (IoCs), threat actors' tactics, techniques and procedures (TTPs), prioritised and filters security alerts and threats, threat intelligence reports, high-level business strategic reports. Intelligence reports intend to meet decision makers' requirements at strategic, operational, tactical, and technical levels. Dissemination and sharing cyber threat intelligence enhance defensive and mitigation strategies for specific risks. Sharing helps the organisation to obtain an efficient situational awareness process and helps cyber risk management.
Evaluation and feedback: This stage also provides continuous feedback about the cyber threat intelligence lifecycle. The feedback determines the quality and usability of the extracted intelligence by the consumer, avoiding requirement gaps, which can be conspicuous once the intelligence report is generated.

2.4.2 Cyber threat intelligence subdomains

There are four different subdomains of threat intelligence [89] [90]:

- Strategic threat intelligence
- Operational threat intelligence
- Tactical threat intelligence
- Technical threat intelligence

Each sub-domain has a different audience, context, collection resources and serves different analytic tools. High-level management of the organisation consumes strategic threat intelligence to make strategic business decisions. This information contains high-level technical information, such as the financial impact of different cyber-attacks. Usually, reports or briefings are the main form of this type of information shared during meetings or high-level conversations. Moreover, because this type contains organisations' strategic business information related to that specific business, sharing such information should stay internal. Analysing this type of intelligence will help to build custom strategic decision making related to a specific business. Analysing this information needs the human effort of collecting and analysing information along with understanding the world's geopolitical situation. For example, if a company is planning to open in a certain country, then they need to know how international events and overseas policies can affect the cybersecurity of the organisation. In addition, they need to know what type of cyberattacks or threat actors the organisation might face in that country.

The second type is tactical threat intelligence which provides more technical information about the attacker tactics, techniques and procedures which have been used to perform an attack. The analysis result will be consumed by incident response teams, security operations managers and administrators. Moreover, this information helps analysts to evaluate different incidents and find the correlation between the activities and attack campaigns. Sharing this information helps to augment the intelligence of the sharing community. The main goal of this type is to get intelligence about the cyber attacker and the tactics they use in order to help protect the organisation and develop mitigation and detection plans.

The third type is operational threat intelligence. This type contains extra information regarding the adversaries' capabilities. It gives more context about the predicted cyber-attacks against the organisation, such as potential risks, attack types, threat actor methodologies. Also, it predicts when the attack will occur. For example, threat actor APT29 [91] tries to gain access through a spear-phishing with self-extracting RAR and uses tools such as TOR, Google Docs, PowerShell scripts to communicate with the compromised systems to control them.

Finally, technical threat intelligence mainly contains compromise indicators from threat actors' tools used in the attack such as IP addresses, URLs and file hashes. This type of information usually comes as machine-readable and automated forms which helps security engineers update endpoint devices and security controls with IoCs feed. This information is shared via open source, public forums, and trusted communities to stop similar attacks automatically.

There are several categories for information sharing architectures, which can be divided into three main architectures based on [92]: centralized, peer to peer and hybrid architecture. In a centralized architecture, the participants share CTI information with the hub where it acts as a repository for intelligence, and then the hub distributes this data to other participants. Distributing this information could be direct to all participants or after performing analysis operations such as aggregation, correlation or adding more context via enrichment of this data. A key advantage of using a centralised architecture is to reduce the cost because it uses standard data formats and protocols. Since the data will go through the central hub, less maintenance and operations are required as they need few connections. The main downside of using this architecture is that a single point of failure might cause a risk of delaying the exchange of information. Another downside is that the centralized hub may raise the motivation for cyberattacks because of the centralisation of a huge amount of sensitive information. Finally, it is a single point of trust, so all community members must trust the centralised hub, and they will be affected in the case where the central hub is not able to keep the sensitive information safe. For example, in the UK the NCSC [93] and its Cyber Security Information Sharing Partnership (CiSP) [94] are considered a centralised information sharing organisation. It is considered the authority for sharing knowledge, threat intelligence and providing a guide on cybersecurity. Also, it provides effective cyber incident management and response against cyber threats to reduce harm to the UK. This body is the interface and builds collaboration between government, industry, SMES and the public to guarantee that the UK is safe online.

In the peer to peer architecture model, members share information directly with one another instead of sharing it with the central repository. This form would be useful when sharing sensitive information about a specific attack with specific peers. The main advantage of this model is that the information is exchanged swiftly between the sharing community members, and there is no single point of failure for targeted attack. On the other hand, if organisations do not support standard formats and protocols, it would be hard for the participants to share effectively. Also, in this type, it is the member's responsibility to enrich each event and perform the analysis operations. Thus, the operations' costs will increase significantly as the number of members increases.

Finally, the hybrid form is a combination between the previous architecture types where sharing decisions will be different based on the context and the situation. For example, an organisation may share more Indicators of Compromise (IoCs) such as hash files and IP addresses using peer-to-peer architecture, while using centralized architecture for sharing enriched events.

There are various sources whereby information can be collected [95] [96] [97]. The first type of sources is internal, such as system logs and network events. The second type is known as Externally Sourced Observables or Feeds such as abuse.ch [98], blocklist.de [99] and MISP (CIRCL) [100]. Also, organisations have open-source intelligence (OSINT) as publicly available sources to be used during data collection phase. An example of OSINT, social media which is one of the key sources to trace information being shared by academics, professionals, or even commercial organisations or research centres. Whois lookup identifies information about the owner of a website, domain name and registers user details. For example, threat actors may use a registered domain to start a social engineering campaign. Other information could be collected through search engines, web services and website analysis.

To achieve cooperative cyberdefense, organisations need to extract the feeds which are related to the cyber-attack and find the correlations with the threat landscape and make the process automated to help the decision-making process. Measuring the feeds' quality and using the proper system controls require a developed standard to help various communities to share and automate cyber threat information. At the same time, the standard should be human-readable as well as machine automated. This will allow cyber threat intelligence analysts to get more context. Also, it allows human processing and decision if needed.

Software vendors and researchers have started to develop and provide models and platforms that help sharing this information, such as "Threat Intelligence Sharing Platforms" (TIPs) [101]. These platforms provide automated support to information sharing and associated analysis. TIPs support organisations to start the processes of collection, processing, analysis, production, and finally dissemination and integration of threat intelligence. There are two main types of TIPs: open source and commercial. These types make it challenging for experts to select which one to implement. Most platforms provide similar operations to aggregate, analyse and support multiple data formats.

For example, Collective Intelligence Framework (CIF) [102] is an open source cyber threat intelligence management system. CIF allows combining information such as IP addresses and URLs from various sources and uses this information for threat identification, detection and mitigation. GOSINT (Cisco, 2017) is an open source threat intelligence platform that can be used for collecting, processing, and analysing indicators of compromise (IoCs). Cisco CSIRT developed this platform with the capabilities to parse indicators from different sources, such as Twitter. It supports searching/sorting, editing and deleting indicators' operations and the ability to add and remove tags of an indicator.

Malware Information Sharing Platform (MISP) [103] is a free and an open source threat intelligence platform for sharing cyber threat intelligence. It supports collecting, storing and processing the relationship between IoCs of cyber-attacks and financial fraud. MISP users created public and closed sharing communities. Each community has different joining conditions [104]. In this platform, vendors created specific formats to support CTI platforms and applications, MISP format is a JSON format to exchange events between MISP instances [100].

Yeti [105] is another open source platform to categorise observables such as geolocation IPs, IoCs, cyber attacker techniques, and information on threats in a centralised repository. It collects observables from various sources such as MISP instances, malware trackers XML feeds and JSON feeds. It provides web API as an interface for machines and web interface for users to integrate with other controls. This platform helps incident responders to skip the “Google the artifact” step of incident response. Also, there are many open source platforms such as MITRE’s Collaborative Research Into Threats (CRITs) [106] and Palo Alto’s MineMeld [107].

There are commercial TIPs available in the market, such as Anomali ThreatStream [108] which supports threats detection by integrating the platform with other security solutions. It collects cybersecurity information from various sources such as commercial and open source intelligence providers, structured and unstructured feeds. IoCs can be sent automatically to other security controls for acting and monitoring. EclecticIQ Platform [109] is also a commercial threat intelligence platform. This platform supports STIX and TAXII standards and provides analyst-friendly graphs with advanced search as well as the ability to integrate with other threat intelligence providers.

Some additional commercial threat intelligence platforms are LookingGlass [110], NC4 Soltra Edge [111], Micro Focus’ Threat Central [112], ThreatConnect [113], ThreatQuotient ThreatQ platform [114] and TruSTAR threat intelligence exchange platform [115]. Both commercial and open source platforms have various limitations [116] based on experts’ views, literature and feedback. Some of these limitations are:

1. The amount of data collected is huge, which makes it hard to generate

intelligence out of it [117].

2. Recently, the focus was on building platforms and data formats and standards, but now it should be more about how to manage this information and reduce the manual work which depends on the analyst [118] [117].

3. Most of the platforms look for tactical indicators and indicators of compromise without always including the context. This limitation might prevent the analysts and the recipients of the information completing the analysis.

4. Most of the existing platforms focus on the collection phase more than the analysis, production and dissemination phases.

5. Only a few platforms support multiple analytical capabilities such as aggregation, advanced searching and visualizations. Many of the existing platforms have limitation in supporting these features [118]. They also often fail to support the business workflow, for example, the ability of users to send a request for information through the threat intelligence platform.

6. Level of trust is related to the users and the platform providers [101]. When sharing cyber threat intelligence information the main trust relationships are between the organisations and the platform provider, which handles the shared information and prevents exposure of confidential information to unauthorized participants. The organisations trust the rest of the participants, and how they will handle the shared data; the platform provider and the participants trust the entity who shared data regarding the reliability and credibility.

7. The platform provider and participants should consider legislation when such big datasets are being processed and shared, such as the GDPR. In this thesis, we have proposed a new framework that would help technical

people when deciding how to share cyber threat intelligence information under the GDPR.

8. Most of the feeds do not record a level of confidence which is related to the quality of the shared information. Therefore, there is a need to measure the quality and confidence of the information from different aspects such as the receiver, the sender and the sharing community [118].

9. Adding time to live property to the feeds is not provided by most of the platforms. This property is critical because the result of this intelligence would be used to prioritise the action during a specific time [118].

10. Threat intelligence platforms do not support a capability to help the organisation calculate the risk of the CTI dataset they are willing to share. In Chapter 4, we propose a new risk assessment model to evaluate the risk of sharing cyber threat intelligence. Our model could be implemented as a component inside the CTI platform.

2.4.3 Cyber threat intelligence standards

In order to build an effective exchange of cyber threat intelligence and use the data correctly for automation, data formats and standards are needed. Many standards have been proposed, and others are still under development for the automated exchange of cyber threat information. These include Cyber Observable eXpression (CybOX™) [119], Structured Threat Information Expression (STIX™) [120], An Open Framework for Sharing Threat Intelligence (OpenIOC), Incident Object Description Exchange Format (IODEF) [13] and Automated eXchange of Indicator Information (TAXII) [121] [101].

Structured Threat Information eXpression (STIX)

Structured Threat Information eXpression (STIX) [120] is an open source language for representing cyber threat information. It was developed in collaboration by a variety of parties under the OASIS umbrella that are 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 [101]. In this thesis, a real-world dataset of cyber incident reports was explored. The MITRE provides a repository which contains STIX incidents used for testing of the STIX schemas [122]. Figure 2 shows the data model of STIX 1.2, that consists of nine main classes with relationships between them.

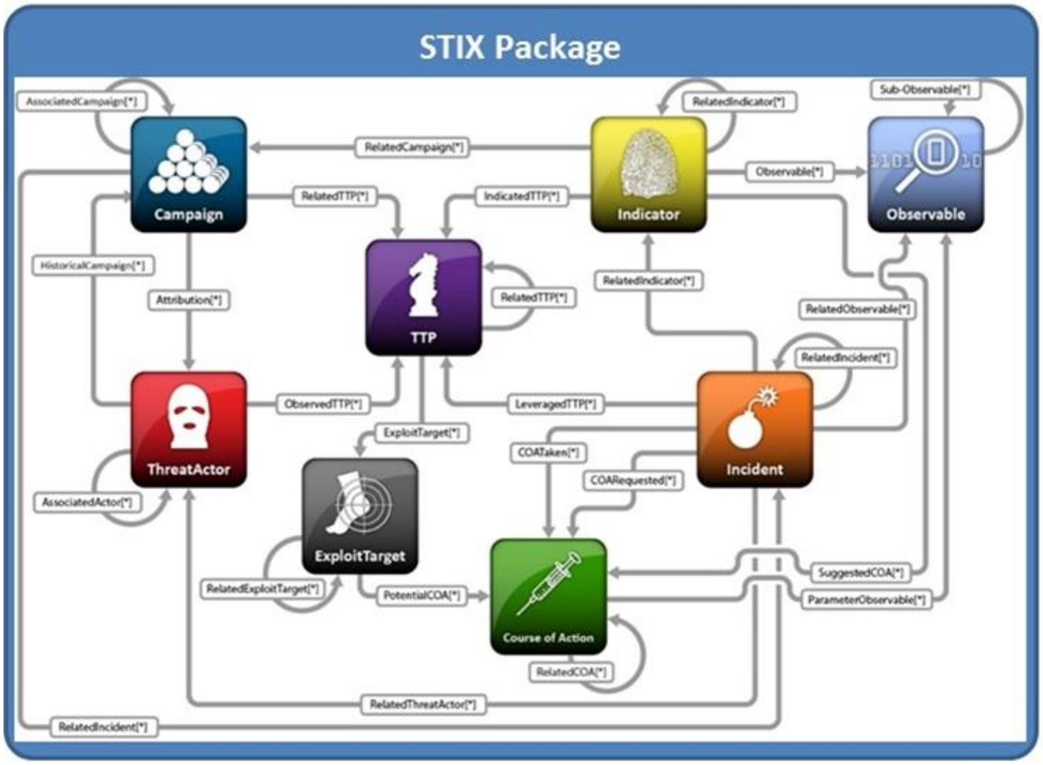


Figure 2: A STIX Package includes the STIX individual component data models [2]

The primary individual component data models of STIX1.2 are [2]:

- **Observable**: STIX Observable describes the properties of what has occurred or what might occur in a cybersecurity event.
- **Indicator**: A STIX Indicator is one of the Observable patterns associated with contextual information. This information describes data that might be related to a cybersecurity event or incident and might be relevant to attacker behaviour.
- **Incident**: A STIX incident is a set of indicators affecting the organisation associated with information found or planned during an incident response process. It consists of properties such as timestamp information, assets affected, impact assessment and related indicators. The STIX incident model

is covered in greater detail in Chapter 3.

- **Tactic, Techniques and Procedures (TTP)**: TTPs describe the attack pattern, techniques, tools, behaviour of threat actors targeting the victim. It contains specific threat actor behaviour such as attack pattern, malware and exploits. Tactics of a threat actor represent how the threat actor operates during different steps of their attack. These tactics include initial access, execution and so on. Each tactic includes various techniques. For example, initial access includes Phishing and Valid Accounts. In order to gain access to the system, an adversary needs to implement the techniques by performing a sequence of actions (procedures) in their attack cycle. For example, the procedure of implementing a social engineering attack is the steps taken by the adversary to target an individual. This attack entails information gathering to identify a target individual, establishing a relationship by writing a convincing email and adding a link or malicious attachment that can bypass the existing antivirus detection and establish command and control of the victim server or workstation.

- **Campaign**: The STIX campaign data model describes one or more instances of cyber threat actors observed via sets of incidents and TTP that intend to attack the organisation.

- **Threat Actor**: The STIX ThreatActor data model represents identification and/or description of the attacker. We might find the same threat actor with multiple names because organisations follow and observe similar activity by different names.

- **Exploit Targets**: The STIX ExploitTargets data model describes vulnerabilities in applications, organisation infrastructure or configurations that target the victim by the TTP of a threat actor.

- **Course of Action**: It describes the potential response and measures for an incident in order to prevent, mitigate or recover from it. As a data model, it consists of related phases of cyber threat management, type of the course of action such as patching or monitoring, the impact of applying this course of action such as high or medium.
- **Report**: The STIX Report data model gives context to the STIX components. It consists of properties such as Title and Time.

The STIX versions 2 and 2.1 have been developed based on STIX 1.X but using JSON instead of XML as a serialisation mechanism. It becomes more lightweight and dynamic with proposing several new objects, such as ‘sighting’ and representing the relationship between the objects via a ‘relationship’ object that can be utilised to link any STIX object [123].

STIX 2.1 represents the cyber threat intelligence by the following objects: STIX Bundle Object and two main categories of objects, STIX Core Objects and STIX Meta Objects. STIX Core Objects have three subtypes: STIX Domain Objects (SDO), STIX Cyber-observable Objects (SCO) and STIX Relationship Objects (SRO).

STIX Meta Objects contain two types: Language Content Objects and Marking Definition Objects; STIX is a connected graph model consisting of nodes and edges. SDO and SCO represent graph nodes, and SRO represents graph edges.

This graph-based language supports analysing related information and allows flexible and agile representations of complex information of CTI.

Each STIX Domain Object consists of properties’ information and related information. We can group some SDOs based on similarity into different categories. For example, Attack Pattern, Infrastructure, and Malware represent types of tactics, techniques, and procedures (TTPs). They describe behaviours and resources that attackers use to perform their attacks and gain access.

For example, with the “Relationship” object we can define a relationship type “uses” to represent the relationship between “Attack Pattern” object and “Tool” object. The data in the example could describe a specific tool such as LOIC (Low Orbit Ion Cannon) [124] which is used to create the behaviour identified in the attack pattern such as a DDOS attack.

pattern

```
{
“type”: “relationship”,
“id”: “relationship–XX”,
“spec_version”: “2.1”,
“created”: “2020-03-06T10:14:22.231Z”,
“modified”: “2020-03-06T10:14:22.231Z”,
“relationship_type”: “uses”,
“source_ref”: “attack-pattern–05”,
“target_ref”: “tool–06”
}
```

Attack Pattern
Uses
Tool

Figure 3: Sample Relationships between SDO

To share STIX reports, we can use a standard called Trusted Automated Exchange of Intelligence Information (TAXII), which defines the technical specification, supporting documentation and the requirements for transporting STIX messages. TAXII is an application layer protocol used to exchange cyber threat information over HTTPS. Thus, it is used to support various protocol bindings and sharing data in various formats [121].

The Incident Object Description Format (IODEF)

The Incident Object Description Format (IODEF) [13] defines a standard for representing cyber incident information using XML schemas to represent multiple classes and data types. It provides a specific set of properties combined with free-text fields for additional unstructured information within the incident data model.The IODEF data model consists of a set of classes to describe an incident. The main classes are Contact, Time, Method, Assessment, History and EventData class. IODEF V2 [125] is an extension of version 1 proposing new objects indicator, threat actor, campaign, and course of action. Furthermore, the property related to time information, techniques used by the attacker, course of action and impact assessment help to provide a detailed representation of the incident.

The Vocabulary for Event Recording and Incident Sharing (VERIS)

The Vocabulary for Event Recording and Incident Sharing (VERIS) [126] is a framework designed to represent cyber incident information based on strategic purpose and risk management. Also, a part of the STIX incident model was derived from the VERIS structure [127]. VERIS depends on a lightweight JSON format. Moreover, it provides explicit definitions for the use of controlled vocabulary or enumerations that enable single value definition for fields, such as type of assets, affected by the incident. A cybersecurity incident is represented by five main components, namely Incident tracking, Victim demographic, Incident description, Discovery and response and Impact assessment. VERIS is more appropriate for reports and briefings since it describes the attacker, defender and the impact information with little information related to the attack.

As a result, we can notice that there are various formats and standards used to represent cyber threat intelligence. This adds challenges and limitations for existing cyber threat intelligence platforms. Much work needs to be done in the direction of enhancing the integration and supporting the exchange of information between different platforms. In the community, there are several efforts to build connectors between existing formats and standards [118]. However, there are challenges for threat intelligence platforms in supporting gathering, exploiting and exchanging of information between various standards and formats. In addition, when converting information between different formats even when upgrading to a newer version, parts of the context and data may be lost [116]. Also, the existence of multiple types of formats and standards are justified due to various purpose of representing and collecting cybersecurity information such as: Yara [128] and sigma [129]. A comparative analysis of incidents reporting standards was based on a set of criteria proposed by [130]. Researchers used three main areas of incident reporting formats evaluation criteria:

- Structural evaluation criteria: based on entity representations which consist of Indicator, Attacker, Attack, Defender and consistent coverage.
- General evaluation criteria: machine-readability, human-readability, unambiguous semantics, interoperability, extensibility, aggregability, practical application and external dependencies.
- Additional evaluation criteria: information such as Licensing terms, Maintenance efforts, Documentation.

The researchers found that the best formats for an automated exchange of cyber incidents information are STIX and STIX2. This result was based on the fact that STIX 1.X and 2.X provide a clear and detailed data model with less ambiguity. STIX introduces extension ability and easy automation because it is machine-readable. Also, STIX is supported by various cybersecurity products, services and user communities (MITRE, 2018). As a summary, this support helps to enhance this standard to enable effective sharing of cyber threat information. In this thesis, we will use STIX 1.X, which described thoroughly in Section 2.4.3.

2.4.4 Information sharing challenges

There are barriers and challenges when sharing information for cyber intelligence analysis. The most obvious of these is the risk associated with disclosing protective capabilities and sensitive information, such as identifiable information and financial loss. Other barriers include business decisions [131], trust between users and platforms providers, different privacy laws, as well as technical issues arising from different platforms and standards [132] [133]. There is the challenge of collecting, processing and managing CTI information, primarily when information is collected from multiple sources in multiple formats, as we mentioned previously. There are various formats and standards that provide free text fields which make it hard to perform an automated analysis [116].

The human aspect is essential in CTI sharing. The process of sharing and dealing with CTI data still needs human users in the process, especially in the identification, remediation and prevention phases [134]. The prioritisation of threats and how to evaluate them differ between organisations where each organisation has a different opinion about the severity of threats [135]. Each organisation will have different types of assets that need to be evaluated and prioritised based on the possible risks and the potential impact on the organisation's business. Thus, they need to identify the possible threats and adversaries that may target them. 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 but want to get benefit from the others ("free-riders"). Also, in general, trust between the sharing partners needs to be established, for example, when they are potential competitors. A straightforward 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 regulation may be a factor that inhibits the sharing of information. Also, one of the critical challenges when sharing cyber threat intelligence is to preserve the confidentiality of individuals and organisations in the sharing community. 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 organisation's cybersecurity strengths and weaknesses, as well as competition sensitive information about business processes. Disclosure of any of this information may raise threats. Threats may be to the organisation's reputation or derived from concerns of intellectual property, business confidentiality or data protection. However, this perceived risk has not been previously analysed in detail. So far, fear of cyber intelligence sharing has been based on a general sense that information in an incident report, for example, would pose some risk if disclosed.

2.4.5 Threat types

In order to identify the cybersecurity and privacy threats associated with sharing CTI, we need to examine the existing cyber threat taxonomies literature.
In [136], the authors propose a taxonomy for cyber-physical threats where they study the attack vectors and the impact on systems and users in the smart home environment. In this taxonomy, they classified attack vector into five main categories, which are: communication medium, supply chain, side channel, sensory channel and control software. Also, they classify the impact on systems into physical impact such as incorrect actuation and cyber impact such as integrity. Finally, they classify the impact on domestic life (DL) into four main categories, which are: emotion regulation and coping, emotional, user experience and direct consequences. In [137] they build a taxonomy used to classify organisational cyber harms based on the risk and impact of cyber-attack. In this taxonomy, they have classified harms into five broad types. First, the primary type is physical or digital harms, such as 'identity theft' and 'Pain'. Economic harms are the second

type, and it includes ‘Reduced profits’ and ‘Disrupted operations’. Also, psychological harms represent the third type and include ‘Confusion’ and ‘Discomfort’. The fourth type is reputational harms, such as ‘Damage public perception’ and ‘Loss of key staff’. Finally, social or societal harms entail ‘Disruption in daily life activities’.

There are other threat taxonomies for grouping threats, for example, Open Threat Taxonomy [138]. Open Threat Taxonomy is an open source taxonomy. It is a well-known description of levels of threats to information systems that organisations may face. The mission of this project is “To maintain a free, community driven, open source taxonomy of potential threats to information systems.”. The authors defined the threats based on the following components: threat providers or agents, threat actions, threat targets and threat consequences. Based on these components, we can characterise the following attack: a threat source such as Lazarus Group [139] performed a threat action such as distributed denial-of-service (DDoS) which led to threat consequences such as availability, confidentiality and integrity. This taxonomy consists of 75 threats action classified into four main categories of threats that could affect the confidentiality, integrity, or availability of information systems such as the following:

- **Physical threats**: Threats with a physical nature that describe actions that could cause destruction of property, loss of property, theft, and harm of information systems.

- **Resource threats**: threats to the resources that are required by the information systems. These types of threats could lead to failures of information systems due to disruption of resources such as water, fuel, electricity required for operations.

- **Personal threats**: Organizations depend on professional personnel to manage information systems. However, when those individual experience disruptions, lack of knowledge, or be wrong, the result can be loss or damage to the secured information systems, for example, skill shortage, social engineering and personnel mistakes.

- **Technical threats**: they are technical threats to information systems. These threats are included when identifying threats executed by a threat agent which could cause damage or loss to an information system.

As part of developing this taxonomy, there was an effort to rank the identified threats and assign a score to each of the identified threat actions. This ranking aims to help the organisation build their risk profile and select controls to stop a specific threat.

The Taxonomy of Operational Cyber Security Risks (TOCSR) [140] was updated in 2014 [141] to follow the security and privacy controls of the 4th version of NIST SP 800-53 [142]. In addition, they tried to create a relationship with other risk frameworks such as The Federal Information Security Management Act of 2002 (FISMA) [143] and OCTAVE [140]. This taxonomy defines operational cybersecurity risks as “operational risks to information and technology assets that have consequences affecting the confidentiality, availability, and integrity of information and information systems”. This taxonomy tries to identify and classify the sources of cybersecurity risk into four main classes based on a business risk perception. The four main classes consist of:

- **Actions of people**: the action or lack of action performed by personnel on purpose or accidentally that affects cybersecurity.

- **Systems and technology failures**: failure of devices and information systems.

- **Failed internal processes**: internal business processes that affect the capacity to establish sustainable cybersecurity, such as process design, modelling, execution and monitoring.
- **External events**: they consist of intractable external problems facing the organisation, such as legal implications, natural disasters and service dependencies.

In this taxonomy, each class consists of subclasses which are described by properties.

Finally, there is the threat taxonomy list from the European Union Agency for Network and Information Security ENISA [116]. ENISA is a centre of expertise in Europe launched in 2004 and based in Greece. It provides support to the European Union (EU) member states and organisations to help better manage cyber risks and meet security policies and regulations. In [144] Launius conducted a review on some of the existing threat taxonomies and found that ENISA's taxonomy contains the most threat actions in the study. Also, ENISA's taxonomy is second in clarity for threat terms and events that are classified under the right class. As a result, ENISA's taxonomy has a high score in the ability to characterise all potential threats to organisations. Thus, in our systematic analysis that follows in Chapter 3, we will use the ENISA threat taxonomy [145] for categorising the threats. ENISA's taxonomy consists of three levels. The top-level categories of this taxonomy include:

- Physical attack (deliberate/ intentional)
- Unintentional damage/loss of information or IT assets
- Disaster (natural, environmental)
- Failures/ Malfunction

- Outages
- Eavesdropping/ Interception/ Hijacking
- Nefarious Activity/ Abuse
- Legal

The next two levels of the taxonomy are **threats** and **threat details**. These detailed levels make it one of the most comprehensive threat taxonomies. This taxonomy focuses on perpetrators' actions that can harm or disrupt information systems and places them into a high-level threat category. This taxonomy defines legal threats in one of the high levels of threat categories. It includes threats of legal or financial penalty due to violation of the law, illegal use of data and court orders.

2.4.6 Legal requirements on cyber information sharing

Cyber information sharing takes place in a legal context which means we have to consider different laws and regulations from different countries. Laws and regulations may both encourage and inhibit aspects of cyber information sharing depending on the country [146]. We get more effective results from sharing CTI data when more participants are involved in the process; thus, encouraging sharing would be ideal if it can be implemented in laws and regulations. In the USA, an executive order, called Improving Critical Infrastructure Cybersecurity, was signed by president Barack Obama [147]. The main goal of this order was to enhance security and resiliency of critical infrastructure by improving the collaboration among federal agencies and voluntary private owners and operators of critical infrastructure. Subsequently, a guideline about sharing cyber threat information which describes what, when, and how to share cyber-threat information was introduced under the Federal Information Security Management Act

(FISMA) [148]. This guideline uses FISMA as a legal baseline of sharing cyber threat information. In the EU, the main relevant laws in this context are the General Data Protection Regulation (GDPR), and the Directive on Security of Network and Information Systems (NIS Directive), both enforced as of May 2018.

General Data Protection Regulation (GDPR)

The General Data Protection Regulation (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 binding to all EU member states [149] [150]. Key terms and concepts related to data protection regulation are the following:

- Data Subject (Article 4 (1)): a natural person about whom data is being collected or processed.
- Personal Data (Article 4 (1)): is "any information relating to an identified or identifiable natural person: one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person". (There are also "special categories" of data, called "sensitive personal data" in previous laws; we will use the term "sensitive information" in this thesis in an informal rather than this legal sense.). This category of data is broader than PII (Personal Identifiable Information), which is personal data that is directly attributed to a specific individual.
- Data Processing (Article 4(2)): the categories of data processing include "any operation or set of operations performed upon personal data or sets of personal data, whether or not by automated means, such as collection,

recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction"

- Data Controller (Article 4(7)): the individual or business who determines the goal of the processing of personal data.

- Supervisory Authority (Article 51): each member state shall provide one or more independent public authorities to be responsible for monitoring the application of GDPR. The goal is to protect the fundamental rights and freedoms of natural per-sons in relation to processing and to facilitate the free flow of personal data within the Union. This role has in most cases been assigned to existing data protection authorities such as the ICO [151] in the UK.

- Personal Data Breach (Article 4(12)): "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data, transmitted, stored or otherwise processed". For example, a hospital could be responsible for a personal data breach if a patient's health information is inappropriately accessed due to a lack of appropriate internal controls.

Any processing of personal data needs legal grounds. Article 6(1) of the GDPR defines the possible legal grounds for data processing as follows:

- After consent of the data subject for one or more specific purposes.

- Processing is necessary for the performance of a contract with the data subject or to take steps to enter into a contract.

- Processing is necessary for compliance with a legal obligation.

- Processing is necessary to protect the vital interests of a data subject or another person.
- Processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller.
- Processing is necessary for the purposes of legitimate interests pursued by the controller or a third party, except where such interests are overridden by the interests, rights or freedoms of the data subject, but this shall not apply to processing carried out by public authorities in the performance of their tasks.

The legitimate interest “must be real and not too vague”. Recital 49 legitimises the processing of personal data for cyber incident sharing, by admitting 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”. Other legitimate interests are processing in order to identify and prevent fraud (Recital 47), or the transmission of personal data within a group of undertakings for internal administrative purposes, including client and employee data (Recital 48) – both these purposes are also related to cyber intelligence sharing. Article (33) introduces mandatory personal data breach notification to the relevant supervisory authority, and to the data subjects when a data breach could cause individual harm. Such notification can also be viewed as a mandatory form of cyber incident information sharing. The GDPR is a main driver for improving cyber security in Europe, 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 4 shows some indicative categories of data in cyber incident reports which are more and less likely to contain personal data under the GDPR (these are "properties" of the STIX 1.2 incident model, which will be explained in more detail in Chapter 3).

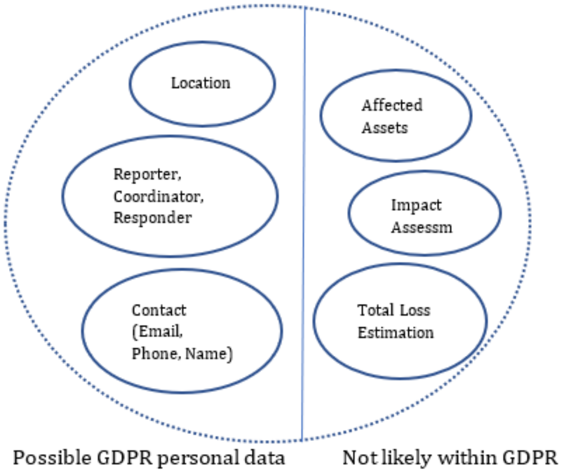

Location
Reporter, Coordinator, Responder
Contact (Email, Phone, Name)
Affected Assets
Impact Assessm
Total Loss Estimation
Possible GDPR personal data
Not likely within GDPR

Figure 4: Example of STIX properties which contain personal information

Talking about the collection and use of personal data, requires transparency in the systems which offers the user the ability to confirm that the information is accurate. Transparency is considered the key to data protection requirements, Articles 12-14 GDPR. Transparency ensures the availability of data before, during and after the processing and can be reconstructed at any time. Thus, transparency should cover what will happen to the data after the processing takes place. Also, transparency is associated with accountability which includes clear documentation covering the source code, technology, responsibilities, privacy policies, notification and the communication with the data subject [116].

Directive on security of network and information systems ("NIS Directive")

The Directive on Security of Network and Information Systems (NIS Directive) is the EU's first piece of cybersecurity regulation. It requires the formation of "Competent Authorities" (CA) which serve and manages cybersecurity within critical infrastructure 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 various countries are expected to share cyber intelligence. Additionally, the CA can audit the OES/DSP's cybersecurity provision, as well as guide them on improvements. Each EU member state needs to define legislation to implement the Directive, including procedures for defining which organisations are OES/DSP as well as "effective, proportionate and dissuasive" penalties for infringement [116]. The NIS Directive assigns ENISA [152] a central role in providing cybersecurity advice and solutions. In the UK, they were able to turn the EU derived NIS Directive into national law that makes it easier for the organisations to be compliant by not enforcing the Directive on an organisation that is already compliant under an equivalent national regulation. The UK was the only country which added a 72 hours deadline for reporting an incident for all OESs and RDSPs.

After two years, the UK has published a review to evaluate how effective the regulations are in terms of costs and benefits [153]. Also, it helps the government to monitor the implementation of the regulations. The main results indicate that after two years of implementing the regulation, it is still early to find out if advantages have materialised. However, organisations started to enhance the security of their networks and information systems as part of enforcing the law; thus, the expectation of applying the law is to reduce the risks and the number of successful cyber attacks against critical infrastructure. However, there is always room for amelioration in the security procedures for the organisation. Still, it needs more time to feel the real impact on the economy and services. For example, the department of health and social care in the UK reported that the WannaCry cyberattack cost the NHS £92m [154]. If, by applying this regulation, we will be able to prevent a similar incident, the benefits of the laws will far exceeded the costs. In a survey with 111 operators of essential services that implemented the NIS Regulation, 61% of operators confirm that the process of recovery from cybersecurity incidents has improved.

The EU Cybersecurity Act

The EU Cybersecurity Act [155] brings new tasks and a permanent mandate for the European Network and Information Systems Agency (ENISA). It improves the EU's ability to respond to cyber-attacks by strengthening ENISA to scale up the collaboration in cybersecurity among EU Member States and EU institutions, agencies, and bodies. The main tasks of ENISA include developing cybersecurity policy for critical infrastructure identified by the NIS directive, such as energy, telecom, and finance. Also, it supports the network of Computer Security Incident Response Teams (CSIRTs) at the EU level in cybersecurity operations and how to handle incidents. It will be providing analysis and technical reports to become the primary source for cybersecurity information from the EU Institutions and bodies. Furthermore, it helps the EU Member states to improve skills and proficiency. Finally, ENISA will conduct marketing related activities within the new Cybersecurity Certification Framework. In addition, the EU cybersecurity act introduces an EU framework for cybersecurity certification. This framework aims to increase trust by providing the technical requirements and rules to evaluate and certify companies' products, processes and services across the EU.

2.5 Risk assessment

Risk is defined in the business world as "the extent to which the outcomes from the corporate strategy of a company may differ from those specified in its corporate objectives, or the extent to which they fail to meet these objectives" [156]. Cybersecurity risks are associated with cybersecurity incidents related to the loss of confidentially, integrity or availability. Risk assessment is one of the essential steps in the risk management process [157]. The risk management process has four steps. The framing risk step involves creating a risk context based on the organisation. During this step, the organisation tries to define organisational risk frame and risk assessment methodology. The second step is assessing risk. This step involves evaluating the risk within the context of the organisation risk frame as defined in the previous step. This step aims to identify threats and vulnerabilities, then to estimate the severity of those threats and finally, the likelihood of threat facing the organisation. The third step is response to a specific risk which includes selecting the right courses of action to align with the organisational risk tolerance and implementing risk response plan based on the chosen course of action. Finally, the last step of the risk management process is monitoring risk. This step involves the follow-up of the current effectiveness of risk responses based on the identified risks. The threat landscape is evolving and at any point that might change the acceptable risk of the organisation. The organisation needs to evaluate the risk responses against threat landscape, business missions and processes and supply chain.

The risk assessment component is the most critical part of the risk management process. There are outstanding risk assessment methodologies to determine the level of risk of security threat, including NIST SP 800-30 [157] which is a framework to help conduct risk assessments of critical infrastructure systems and organizations. This framework allows senior management to select the course of action in response to specific threats. In the NIST framework, the risk assessment methodology consists of four phases. It starts with the risk assessment process, which explains the process of evaluating information security risk. The second step is risk modelling which describes risk factors such as threat source, threat event, severity and likelihood, and defines the relationship between the factors. The third step is the assessment approach, such as quantitative or qualitative assessment. Finally, the analysis approach is determined by the organisation to decide how to combine and analyse threat factors. The analysis approach can be classified as threat-oriented, asset/impact-oriented or vulnerability-oriented.

Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) [158] focus on identifying vulnerabilities that exist in the organization's structure and implementing security strategies and plans. The OCTAVE methodology consists of three phases. The first phase includes prioritising the existing assets in the organisation based on threat profiles. The second phase includes identifying the security level of the critical assets based on the organisations' infrastructure vulnerabilities and possible attacks. Finally, the third phase includes the result of selecting critical assets and identifying their threat profiles. An evaluation of the associated risk of each critical asset is conducted to respond against any possible risk.

Privacy risk assessment is quite close to security risk assessment. In the NIST, privacy risk assessment is defined as "A privacy risk management sub-process for identifying and evaluating specific privacy risks". Privacy risks are linked to privacy events related to data processing. A privacy event is defined as "The occurrence or potential occurrence of problematic data actions". Moreover, a Data Action is defined as "system/product/service data life cycle operation, including, but not limited to collection, retention, logging, generation, transformation, use, disclosure, sharing, transmission, and disposal". Data processing is defined as "The collective set of data actions". Thus, both cybersecurity and privacy risk assessment are connected when cybersecurity incidents are occurring from privacy events [3]. For example, consider installing smart meters and smart appliances as part of the Smart Grid. The ability of these meters to collect, process and manage detailed information about energy use can be conducive to identifying details of individuals' behaviours and their daily life inside their houses. The smart meters are working as planned, but the data processing could suggest that people are under surveillance. Figure 5 demonstrates the overlap and relationship between privacy risks and cybersecurity risks. NIST defines three privacy risks factors to be assessed and combined to get the risk score. The three risks factors are problematic data action, likelihood and impact.

Also, there is a relationship between privacy risks and organisational risks by linking the privacy risks to the defined organisational impacts such as legal penalties and loss of reputation. Besides, the privacy risk assessment involves additional contextual descriptions and an extra level of granularity of the risk level. Using the defined level as a result of the privacy risk assessment may give an impression that it is always a fuzzy decision; thus the result will not be as informative as necessary [159]. Even in the development life cycle, we can find security by design integrated in every step to build secure software. This integration will prevent an attacker from exploiting design flaw. In this aspect, threat modelling such as STRIDE [160] plays a significant part in finding system security threats.

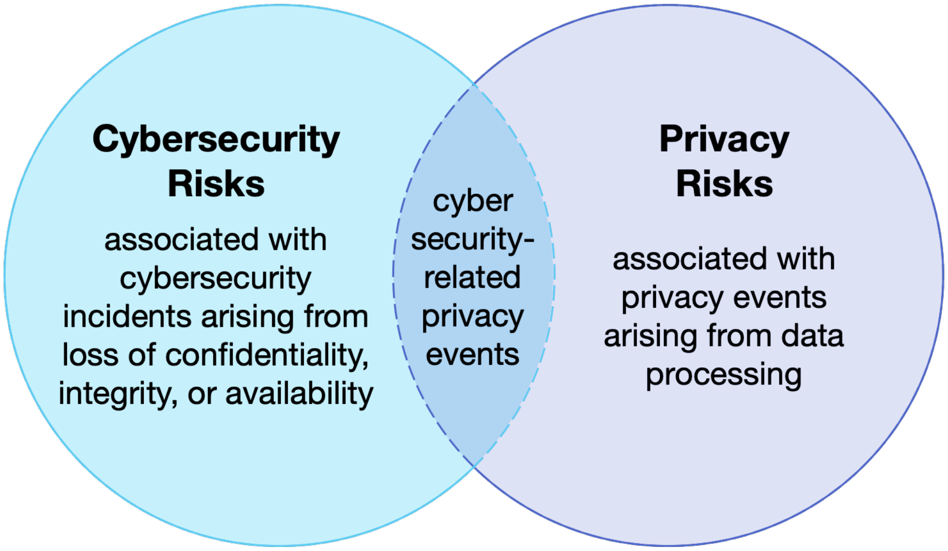

Cybersecurity Risks
associated with cybersecurity incidents arising from loss of confidentiality, integrity, or availability
cyber security-related privacy events
Privacy Risks
associated with privacy events arising from data processing

Figure 5: Cybersecurity and Privacy Risk Relationship [3]

To build a privacy threat modelling that could be associated with the software life cycle, [161] proposed the LINDDUN methodology, inspired by STRIDE, to help developers to protect software architecture against privacy threats. LINDDUN helps to manage seven privacy threat categories: Linkability, Identifiability, Non-repudiation, Detectability, Disclosure of information, Unawareness and Non-compliance. The LINDDUN methodology consists of three key phases: model the system, elicit threats and manage threats. However, LINDDUN allows the analyst to choose the risk assessment approaches as they are not restricted to a specific risk assessment model.

There are various ways to assess risk, including quantitative, qualitative, or semi-quantitative. Quantitative risk assessment uses mathematical methods and rules. Here, numbers represent information; for example, a numerical value of 1 is assigned to the high probability of a specific attack that could occur. Understanding the context and explaining the constraints helps to assign the numbers in a meaningful way; thus, the meaning of the quantitative results would be more precise. However, in some cases, the results need additional justifications and clarifications to understand what the numerical results represent. For example, before sharing any CTI dataset, the owner may ask if the risk assessments results are reliable based on the assumptions used in the calculations. Privacy risk can be evaluated by using various privacy risk metrics to measure impact and likelihood and combine them to get the risk score [159]. They find that different privacy risk metrics could be defined to measure impact and likelihood.

On the other hand, qualitative risk assessment is based on applying non-numerical methods according to levels such as low, medium and high. This type of assessment has a limited number of results which make it more comprehensible to decision-makers. Each value should be defined clearly and categorised by a clear description and an example. Without a clear description, experts may rely on their experience and opinion which might provide different assessment results.

Finally, semi-quantitative risk assessment combines rules and methods for evaluating the risk based on numeric values and levels. For example, the range between 1 and 10 can easily be converted into qualitative expressions that help risk communications for decision makers, for example, a score of 10 can be represented as extremely high. The role of expert judgment in assigning values in the semi-quantitative risk is more evident than in a purely quantitative approach. Moreover, if the scales or sets of levels provide sufficient granularity, relative prioritization among results is better supported than in a purely qualitative approach. In this type of risk assessment, all ranges and values need to be explained and defined by clear description and examples. Semi-quantitative assessments use various methods or rules for evaluating risk based on levels, scales or numeric values that are meaningful in the context. For example, a score of 90 for a CTI dataset can represent a very high risk. The role of experts' judgment still exists and, as with qualitative and quantitative models, each numeric value and range needs to be defined and explained.

2.6 Related work

2.6.1 Risks of sharing cyber threat information

In [148] Johnson looked at the threat associated with sharing sensitive information and financial transition between financial firms. Johnson identifies the connections between the number of leaked documents and the number of threats and vulnerabilities. In that study, they focus on peer-to-peer-file-sharing networks, especially between employees, therefore reducing the risk of the disclosure which reduces the possible threat activity arising from exploiting the leaks. Accidental disclosure of sensitive information represents one of the primary information risks against businesses. Existing tools and technologies for sharing sensitive information create various security risks for these businesses.

Threat actors can apply queries that can be used to extract data from the organisation's files. For example, they might find that John, an employee, is using Microsoft Office 2007 to create sensitive files which leaked accidentally. This risk could be reduced by proposing file-name conventions and enforcing its new policies. However, they discussed only the risks to business without covering the cybersecurity threats that might occur as a result of that sharing.

The authors in [162] focus their study on two factors: the willingness to share cyber threat information and the usage of cyber threat information. They used a survey to assess the privacy risk via ordinal probability range and nominal data types in factorial vignettes survey. They classified the data into different categories which are high-usage, low-risk versus those that are low-usage, high-risk. The list of properties includes Passwords, Usernames, Keylogging data, E-mails, Chat history, operation system information and other properties. For example, there is a high usage of the "Usernames" property in the CTI dataset, but on the other hand, organisations are unwilling to share this property with others due to the high risk level score. However, organisations are more willing to share information such as IP addresses and specific network information. Even for less sensitive information, cybersecurity professionals are asking for applying secure sharing procedures, such as access control and encryption. For sensitive information, cybersecurity professionals stress using data minimisation techniques, such as anonymisation.

In this study, the authors only classified a specific list of properties that might exist in the CTI reports based on the level of willingness to share. There was not a precise analysis of the associated threats of sharing properties that the experts are less willing to share.

Much work has been done in defining principles for sharing cybersecurity information. [163] defined three principles that indicate sharing security information within and between organisations; the principles are Least Disclosure, Qualitative Evaluation, and Forward Progress. The purpose of these principles is to decrease privacy risks in cybersecurity data sharing.

First, the Principle of Least Disclosure: sharing the minimum information within or between organisations to reduce the risk of sharing. The associated corollaries to this principle are internal disclosure in the collection phase, privacy balance to choose carefully the trade-off between utility and privacy, and the final corollary is Inquiry-Specific Release by giving the access to the minimal amount of information based on approved specific uses. There are many approaches to achieve this principle, such as anonymisation and Minimal Requisite Fidelity. In [69] researchers proposed a model for collaborative information analysis systems. The model addresses the trade-off between privacy leakage and utility to address privacy concerns. This model aims to select the best privacy preserving techniques to optimise the trade-off between privacy and utility.

Second, the Principle of Qualitative Evaluation combines technical and legal constraints. Without implementing legal constraints, we might not be able to share information, and at the same time, we cannot rely only on technical methods for applying privacy. In [164], the authors propose a model for sharing cybersecurity information by using a ledger model to store all transactions about sharing CTI datasets and smart contracts by using blockchain technology to enable secure sharing and collaboration. However, it focuses only on a specific issue about sharing CTI datasets. Finally, the Principle of Forward Progress stipulates that organisations should not stop sharing information under the pretext of legal requirements or safety reasons because that will prevent the benefit of sharing and finding solutions quickly.

The most relevant work targeting privacy preserving techniques for cybersecurity information sharing is [165]. They proposed PRACIS, a scheme that guarantees private data forwarding and aggregation for problems related to cybersecurity information sharing networks. They obtained these goals by combining homomorphic encryption primitives and standard format-preserving encryption where the output of the encryption operation keeps the same format as the input. The proposed scheme leverages the STIX standard with the ability to be integrated with current STIX-based message brokering middleware.

The explored studies do not consider legal aspects, and the associated risk when selecting the actions to be applied to the data to mitigate privacy and confidentiality threats of cybersecurity information sharing. Also, they did not take into account ways of retaining the high utility of the shared data.

2.6.2 Risk assessment of sharing cyber threat information

In [166], the authors addressed the types of information that could be shared between SMEs while addressing the risk of disclosure cyber-attack scenarios. However, the study was limited to SMEs and a small size sample with specific security metrics which could be different in various business scenarios. In our work, we evaluate the risk and propose a more general model, not related to specific businesses, for evaluating the risk of sharing CTI datasets.

In [167] the authors proposed a cybersecurity risk model using a Bayesian network model for the nuclear reactor protection system (RPS), they then apply the analytical result to an event tree model. In their model, they only focused on four cyber threats and six mitigation measures for the design specification of an RPS. This evaluation was only on the network layers and did not cover other types of possible threats.

In [168] the authors proposed a quantitative asset and vulnerability centric cyber security risk assessment methodology for IT systems. They defined and extended metrics based on Common Vulnerability Scoring System (CVSS) and presented a formula for computation and aggregation. The work focused only on the Common Vulnerabilities and Exposures (CVEs) without considering the impact of other factors. Also, the calculation was based on the defined CVSS list without including zero-day attacks. The model did not consider the threat actor and the attack vector, as the focus was only on the individual asset and the vulnerabilities of the assets in the the system design. They proposed a base risk assessment model and an attack graph-based risk assessment model.

In [169] the authors propose a model to evaluate the correlation between disclosed security risk factors and future security breach announcements reported in the media. They used text-mining on the reports to enhance and enrich the classification method. The results show that including mitigation steps to the disclosed security risk factors would minimise the likelihood of future breach announcements. They investigate how the market infers the context of security risk factors in annual reports. Therefore, they develop a decision tree model, which categorises the occurrence of future security breaches according to the textual contents of the disclosed security risk factors. They claim the model can accurately associate disclosure features with breach announcements about 77% of the time. The results indicate that the disclosed security risk factors with risk-mitigation themes are unlikely to be pertinent to future breach announcements. They also examine how the market analyses the nature of information security risk factors in annual reports.

In this work, they focused on the disclosed reports of the firms. Also, there was a limitation in the number of these reports, in addition to focusing on the market and financial response without going into the details of cybersecurity risks.

In [21] the authors looked at the effect of sharing vulnerabilities of an ICT system on responsible market and disclosure policies. They found that sharing vulnerabilities with the public would immediately increase the probability of cyber-attack. This sharing gives a road map to attackers to gain access and attack the systems.

This work only considered the disclosure of vulnerabilities which is a part of CTI datasets. Also, they did not investigate legal or technical threats.

In [170], researchers propose a risk assessment and optimisation model to extend the standard risk assessment process and find the balance between the existing network vulnerabilities and financial investments.

In [171] the authors proposed an architecture to compute a privacy risk value of cyber threat information extracted from a STIX report. They build a survey to collect data by using factorial vignette and multi-level modelling.

However, these methods do not adopt a quantitative approach for risk evaluation when sharing CTI datasets, such as the one presented in this thesis. In this thesis, I propose a new model to compute risk by identifying threats, severity and probability of sharing CTI information, which will be described in more detail in Chapter 4.

2.6.3 Sharing cyber threat information under laws and regulations

Many papers have addressed issues related to terms and rules extracted from regulations and policies for protecting personal data. In [172] the authors converted the precursor of the GDPR, the 1995 EU Data Protection Directive [173] into executable rules to support access control policies. The authors presented a system to automate legal access control policy to make an automated decision concerning authorization rights and obligations based on the related legal requirements. In [174] the authors developed a specialised tool for privacy control based on the GDPR to share sensitive research datasets. They used DataTags to categorise datasets. Data-Tags assign a label to a dataset. Each DataTag may contain human-readable and machine-actionable rules. Thus, a dataset will be assigned to a specific label after conducting a series of questions based on defined assertions within a particular context. After assigning a label to the dataset, it will be possible to apply the associated machine-actionable actions to the dataset or building a custom data sharing agreement to be compliant with the human-readable rules. Thus, the authors defined the security measures of the data tags levels based on the DANS EASY repository [175]. The authors focused on datasets managed by researchers in a general context. In [176] [177] the authors extracted data access rights from a legal test of the US Health Insurance Portability and Accountability Act (HIPAA). They used an ontology to classify legal rules of privacy requirements from regulations to give a decision to grant or deny the access right.

In [178], researchers proposed a privacy by design solution to exchange cybersecurity incident information between CSIRTs. This solution focused only on sharing information between closed user circles such as the CSIRTs. The authors aimed to illustrate the legal requirements about sharing CTI datasets which contain personal information between the CSIRTs without giving a systematic way to help the CTI datasets manager to check the legality of sharing such information.

Previous research findings into the legal grounds of sharing CTI information under the GDPR have been inconsistent [179] in comparison with our justification for sharing CTI. They examined sharing under the legal bases of legitimate interest or public interest. As argued by the researcher, they claimed that the legal grounds for sharing CTI can be justified under the public interest under Article 6 (1)(e) of the GDPR, and the notification requirements of Article 14 makes relying on the legitimate interest unjustifiable. A full discussion regarding the possible requirements for sharing CTI datasets will be described in more detail in Chapter 5. In our work, we aim to build a set of sharing requirements that CTI datasets managers will check to provide a decision about sharing CTI dataset(s) under the GDPR.

Chapter 3

Risks of Sharing Cyber Incident Information

In this chapter[1], we present 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. This chapter is divided into the following sections. Section 3.1 describes the methods used for threat analysis. Section 3.2 discusses the analysis of disclosing cybersecurity incident information in the STIX incident model with its key findings. Section 3.3 provides an evaluation of

[1] This chapter is based on the conference paper "Risks of Sharing Cyber Incident Information" [180]

other standards of sharing cyber threat information. Section 3.4 summarises this chapter.

3.1 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 3.1.1 below. We then point out threats, using a taxonomy described in Section 3.1.2. Finally, we assess the severity of privacy and security threats according to a methodology described in Section 3.1.3.

3.1.1 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 anonymisation and de-identification methods [34] [35] which have been described in more detail in Section 2.3. The attributes' types are [181]:

- *Identifier attributes* include information used to identify an individual such as full name, driver license, and social security number. We extend this to the identification of organisations as well as individuals.

- *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.

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.

3.1.2 Threat taxonomy

In our systematic analysis that follows, we will use the threat taxonomy from ENISA [145] for categorizing the threats and breaking down attacks in terms of how they accomplished. This and alternative choices were described in Section 2.4.5. The high level categories of this taxonomy are:

- Physical attack (deliberate/ intentional): This category covers physical threats such as “Fraud” and “Sabotage”. Physical access to the devices could give the attacker the possibility to establish malicious activities.

- Disaster (natural, environmental): This category covers threats which could damage information assets due to natural or environmental disaster for example, “natural earthquakes”, “floods”, “fire” and so on.

- Failures/ Malfunction: this category covers threats which could cause a failure of IT supporting infrastructure such as failure of main supply caused by the failure of cooling infrastructure.

- Outages: this category covers threats that rely on losing resources such as loss of electric power to operate an IT infrastructure. The reason is likely an external factor such as large loss of power in an area due to a fault in underground power cables.

- Eavesdropping/ Interception/ Hijacking: this category covers threats that depend on changes of communication between two entities. These attacks do not need to install extra tools on the victim machine, for example man in the middle/session hijacking attack.

- Nefarious Activity/ Abuse: this category covers threats that need additional steps by installing tools or software on the victim's machine such as protocol exploitation or spoofing.
- Legal: this category covers threats related to legal or financial penalty caused by the existing legislation.
- Unintentional damage/loss of information or IT assets: this category covers threats of human mistakes such as sharing information by mistake or unintentional change of data in information system.

We have considered only threats relevant and associated with disclosing cyber incident information. Table 9 shows the list of threats. We will use the value of the ID column instead of the threat column's value in the analysis and the mapping tables.

3.1.3 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) [182]. 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

ID	Threat
T1	Social Engineering
T2	Loss of (integrity of) sensitive information
T3	Failure to meet contractual requirements
T4	Violation of laws or regulations / Breach of legislation
T5	Compromising confidential information (data breaches)
T6	Failure of business processes
T7	Identity theft (Identity Fraud/ Account)
T8	Unauthorized activities
T9	Targeted attacks (APTs etc.)
T10	Unauthorized physical access / Unauthorised entry to premises
T11	Terrorists attack
T12	Loss of reputation
T13	Manipulation of information
T14	Misuse of information/ information systems
T15	Judiciary decisions/court orders
T16	Man in the middle/ Session hijacking
T17	Generation and use of rogue certificates
T18	Abuse of authorizations
T19	Information leakage/sharing due to human error
T20	Start or Failure or disruption of main supply
T21	Failure or disruption of service providers (supply chain)
T22	Denial of service
T23	Malicious code/ software/ activity
T24	Abuse of Information Leakage
T25	Abuse of vulnerabilities, 0-day vulnerabilities
T26	Brute force

Table 9: Threat List

individual, i.e. the level of identification. Table 10 shows the description of the scores for this on a 1-4 scale, as taken from [182].

The prejudicial effects value of each threat is also scored on a 1-4 scale as given in [182]. Table 11 describes this.

Finally, The CNIL standard [182] computes the severity value by adding the

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

Table 10: Level of Identification

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

Table 11: Prejudicial Effects

level of identifiability and prejudicial effects of potential impacts values obtained and translates that into a risk severity scale as given in Table 12. We record the resulting severity as PS (Privacy Severity) for each risk in our analysis.

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

Table 12: Severity Value

As indicated above, we have generalized this method to also apply to cyber security risks, yielding a Cybersecurity Severity (CSS) score. For this, we use Table 10 and Table 12 unchanged, and instead of Table 11 we use the very similarly constructed Table 13 to score the ease of exploiting cybersecurity information.

Score	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

Table 13: Ease of Exploitation

3.2 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.

3.2.1 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 14 shows an example of cyber incident information report.Table 15 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.

Property	**Value**
TTP Malware Type	Capture Stored Data, Remote Access Trojan
Indicator Name	File hash for malicious malware
Indicator Description	This file hash indicates that a sample of malware alpha is present.
Indicator Value	Hashes.‘SHA-256’= ‘ef537f25c895bfa7jfdhfjns73748hdfjkk5d89fjfer8fjkdndkjn7yfb6c’ Windows-registry-key:= “HKEY_LOCAL_MACHINE\\SYSTEM\\ControlSet001\\Services\\MSADL3”
Vulnerability	CVE-2009-3129, CVE-2008-4250, CVE-2010-3333, CVE-2012-0158, CVE-2011-3544
Incident Title	Incident associated with CyberA campaign. The malware was designed to steal encrypted files - and was even able to recover files that had been deleted.
Date	2012-01-01T00:00:00
Reporter Name	Alex John
Reporter Email Address	alex@pro-it.com
Reporter Address	US-LA
Victim Name	CyberA / The CEO Device
Victim sector	Financial sector
Victim Device	IP address: 146.227.239.19
Victim Email Address	cybera@cyber.com / ceo-cybera@cyber.com
Victim Address	CyberA Ltd, IT Department, LONDON, W5 5YZ
Affected Assets Type	Desktop, Mobile phone, Router, Server, Person
Affected Assets Property	Confidentiality (Classified, Internal, Credentials, Secrets, System) Integrity (Software installation, Modify configuration, Alter behaviour)
Incident Status	Not solved
Total loss	£ 65,000

Table 14: Example of Cyber Incident Information Report

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.

Property	**CT**	**IFT**	**S**	**I**	**PI**	**Threat**	**PS**	**CSS**	**Justification**
version	✗	✗	✗	✗	✗	N/A	0	0	This refers to the report, not to the incident
Reporter	✓	✓	✗	✓	✓	T1, T7, T8 T10, T26	4+2	2+2	The identity of the reporter can be revealed
Description	✓	✓	✓	✓	✗	T1, T4, T5 T19, T20, T21 T24, T25	2+2	2+2	Free text field which is likely to refer to particular business information, and may contain sensitive and identifying information
Security_Compromise	✗	✗	✓	✗	✗	T5, T9, T12 T22, T26	0	3+3	Identifies whether critical information was leaked, and it can be sensitive
COA_Requested	✓	✓	✓	✓	✓	T1, T5, T9 T16, T17, T24	2+2	2+2	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.
Related_Indicators	✓	✓	✓	✓	✓	T1, T5, T9, T12, T22, T24, T25, T26	2+2	2+2	This can refer to specific information about the incident and adversary Tactics, Techniques, and Procedures (TTPs). May contain identifying information about the adversary

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 (Level of Identification + Prejudicial Effects), (CSS) Cybersecurity Severity (Ease of Exploitation + Prejudicial Effects).
For the values of the properties, ‘✓’ denotes ‘yes’, ‘✗’ denotes ‘No’, ‘*’ denotes ‘It depends’.

Table 15: Analysis Sample for Class “IncidentType”

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. This is also an indication of a possible privacy risks and consequently a data breach based on legal risks.

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 3.1.3. 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 12. 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.

3.2.2 Analysis sample

The full analysis is given in Table 3.4 at the end of this chapter. Here we explain a sample of this information in full detail. Table 15 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. Furthermore, Table 3.4 gives full analysis of each property of the IncidentType class of the STIX incident model.

3.2.3 Severity results

We have computed the severity values for each property of the STIX incident model based on the method proposed in Section 3.1.3. In particular, Figure 6 shows the cybersecurity severity results for the first level properties of the STIX incident model. Figure 7 shows the privacy severity results for the first level properties of the STIX incident model.

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

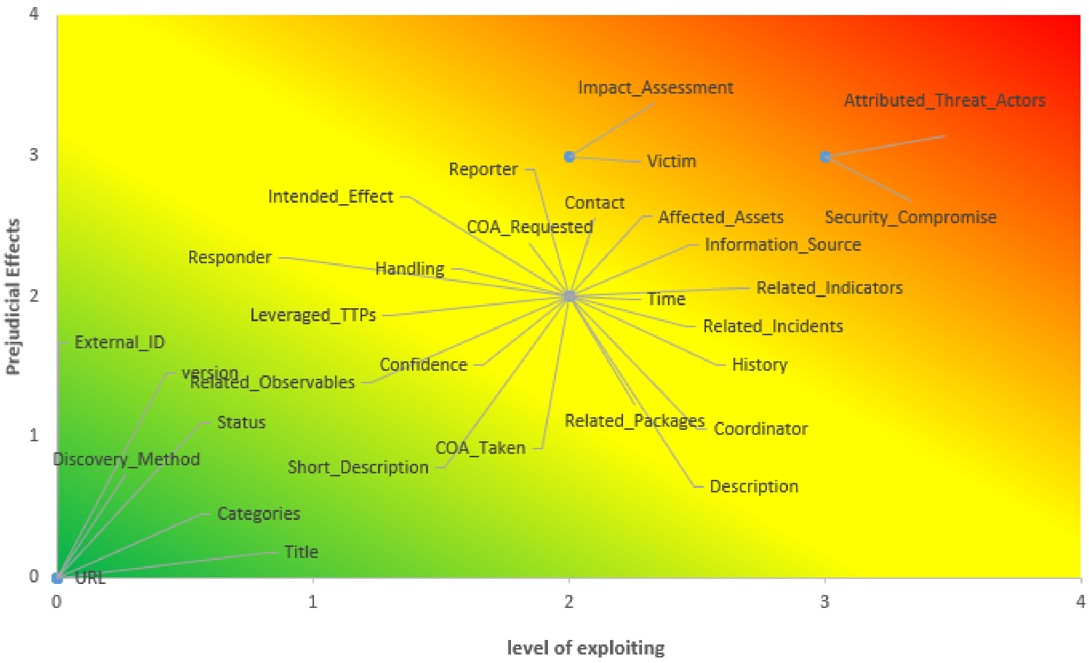


Figure 6: Cybersecurity severity results

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.

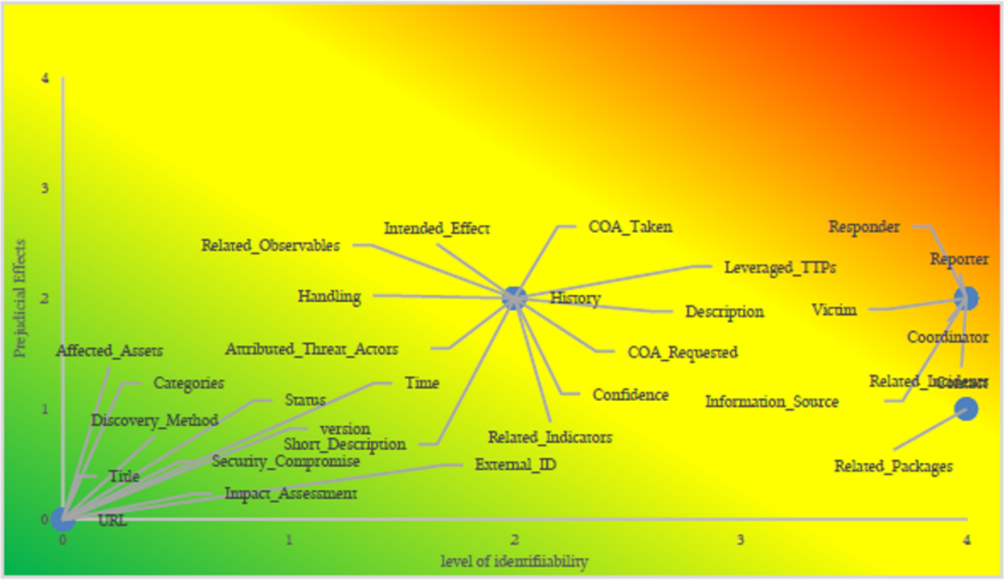


Figure 7: Privacy severity results

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 [183]. This attack started by weaponising the network with BlackEnergy malware using spear-phishing attacks, then hijacking SCADA systems, and remotely controlling electricity substations.

Table 16 provides examples of attack vectors associated with STIX incident properties that the adversaries could use to have an initial access within a system or a network. Most of the attacks happen in one or more steps. Cells in columns “Initial Access Attack” and “Description” are taken from the MITRE ATT&CK framework [184]. The values in column “STIX incident Report Property” are our proposals as to which attributes might be used to initiate the attack.

The values in column “Attacks example/ Threat actor groups” are real-world example of successful attacks or threat actor groups targeting critical infrastructure. As an illustration, the adversary first investigates the intended victim to gather necessary background information so information such as “Description”, “Short_Description”, “Reporter”, “Responder”, “Coordinator”, “Victim”, “Contact”, “History”, “Information_Source” properties could be very useful for this step. Then, the adversary would be able to take advantage of this information to choose a specific individuals or entities and try to exploit a vulnerability on the victim’s system. For example, APT19 [185] sent spearphishing emails containing malicious attachments in RTF and XLSM formats to deliver and execute initial exploits. This threat actor group targeted at least seven global law and investment firms.

Initial Access Attack [184]	**Description [184]**	**STIX incident Report Property**	**Attacks example / Threat actor groups**
Exploit Public-Facing Application	The use of software, data, or commands to take advantage of a weakness in an Internet-facing computer system or program in order to cause unintended or unanticipated behaviour.	Description, Short_Description, COA_Taken, Related_Indicators, Related_Observables, Leveraged_TTPs, Related_Incidents, Security_Compromise, Discovery_Method, COA_Requested, Affected_Assets	CVE-2016-6662 [186], CVE-2014-7169 [187]
Hardware Additions	Computer accessories, computers, or networking hardware may be introduced into a system as a vector to gain execution. While public references of usage by APT groups are scarce, many penetration testers leverage hardware additions for initial access.	Description, Short_Description, COA_Taken, Related_Indicators, Related_Observables, Leveraged_TTPs, Related_Incidents, Security_Compromise, Discovery_Method, COA_Requested, Affected_Assets	Passive network tapping [188], Keystroke injection [189], adding new wireless access to an existing network [190]
Spearphishing Attachment/ Link/ via Service	All forms of spearphishing are electronically delivered social engineering targeted at a specific individual, company, or industry.	Description, Short_Description, Reporter, Responder, Victim, Coordinato, Contact, History, Information_Source	APT19 [185], APT28 [191], APT29 [192], APT32 [193], APT37 [194]
Supply Chain Compromise	Supply chain compromise is the manipulation of products or product delivery mechanisms prior to receipt by a final consumer for the purpose of data or system compromise.	Description, Short_Description, COA_Taken, Security_Compromise, Related_Incidents,COA_Requested, Affected_Assets	CCBkdr [195], Elderwood [196], Smoke Loader [197]
Valid Accounts	Adversaries may steal the credentials of a specific user or service account using Credential Access techniques or capture credentials earlier in their reconnaissance process through social engineering for means of gaining Initial Access.	Description, Short_Description, Reporter, Responder, Coordinator, Victim, Contact, History, Information_Source.	APT28 [191], APT3 [198], APT32 [193], Carbanak [199], Cobalt Strike [200].

Table 16: Examples of attack methods associated with STIX incident properties

3.2.4 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 [201]. 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 [202], 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) [203]. 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) [204]. In general, privacy risks only materialise when a sensitive feature is revealed about an identified actor but the identifying and sensitive features could occur in different STIX properties. As an extreme example, for financial damages, strictly speaking 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 3.2.3.

3.2.5 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 1.2 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.

3.3 Extending to other standards of sharing cyber threat information

3.3.1 From STIX 1.2 to STIX 2

In this section, we discuss the impact of the transition to STIX 2 [123], which is promoted by OASIS [205], on our analysis of STIX 1.2 when sharing cyber threat information. There are two main differences between STIX 1.2 and STIX 2, as follows [206]:

1. STIX 2 is using JSON language in the implementation phase, while STIX 1.2 was defined by using XML.
2. Many embedded properties in STIX 1.2 became objects in the top-level in STIX 2 structure. These objects are called STIX Domain Objects (SDOs) connected by defined STIX Relationships Objects (SROs).

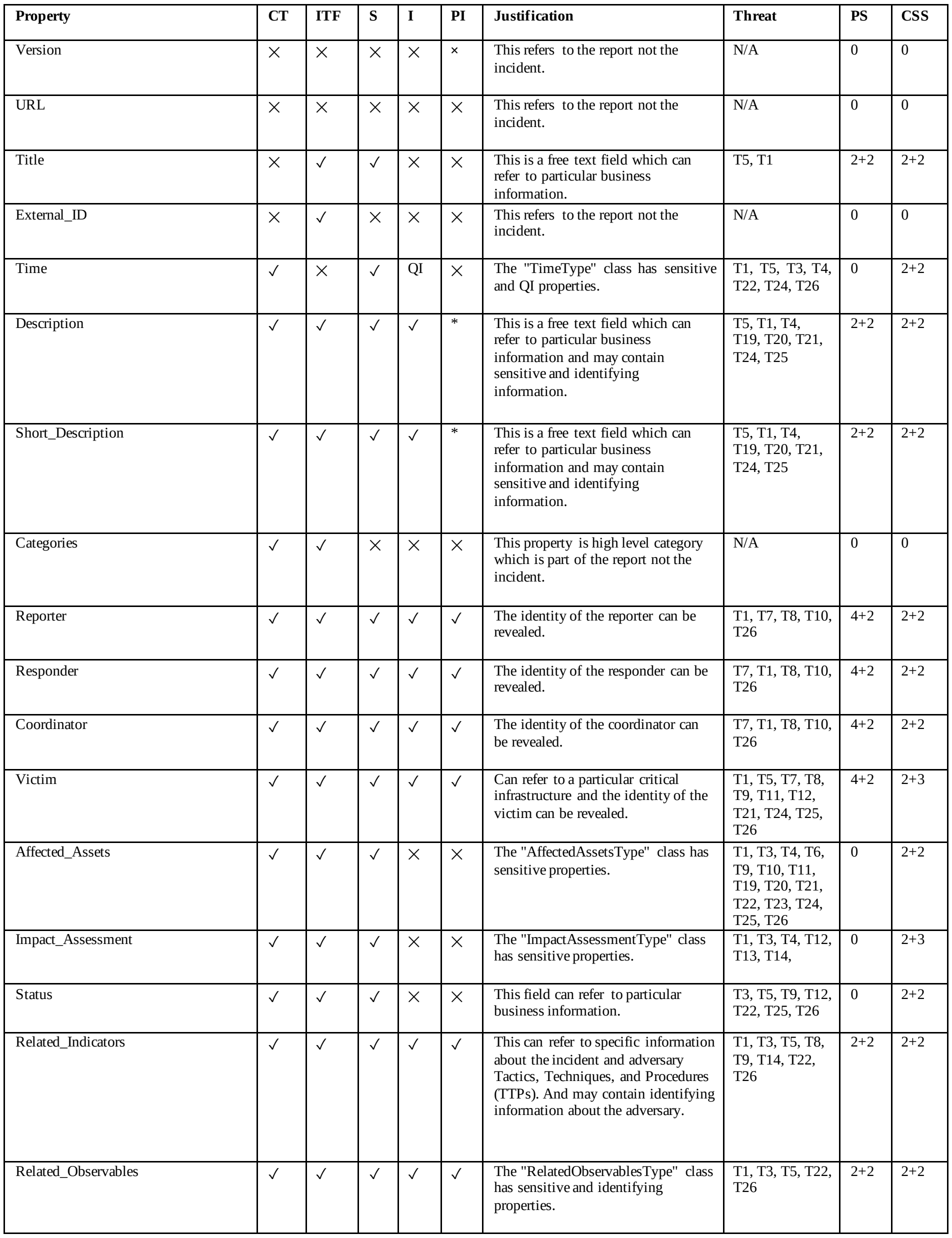
STIX 2 currently defines twelve STIX Domain Objects (SDO), which are Attack Pattern, Campaign, Course of Action, Identity, Indicator, Intrusion Set, Malware, Observed Data, Report, Threat Actor, Tool, and Vulnerability. The incident object has not been developed yet but it is intended to be included in STIX 2.1 [206]. However, many inherited classes in STIX 1.2 are defined as new objects in STIX 2, such as "Course of Action" and "Identity" objects. The "Course of Action" object contains the "Description" property. "Description" property type is a free text field that describes the actions to prevent or respond to an attack. As we have mentioned earlier any free text field might contain sensitive information, hence any exposure of this information would be associated with multiple threats. Another embedded property in STIX 1.2 incident model that was defined as a new object in STIX 2 is the "Identity" object. The Identity object consists of many properties such as name, description, and contact_information. This information is sensitive because it can refer to the identity of the victim. Consequently, the disclosure of this information can lead to threats, such as, loss of reputation and spear phishing attack.

3.3.2 STIX and IODEF

Besides STIX 1.2 and STIX 2 for describing threat intelligence, we have looked at Incident Object Description and Exchange Format (IODEF) version 2 [125], which was released in November 2016. It uses XML to represent and share cyber incident information between Computer Security Incident Response Teams (CSIRTs). Similar to STIX 1.2, IODEF consists of a hierarchy of many classes and sub classes used to describe Assessment, Method such as the attacker techniques, contact information such as email addresses and phone numbers, and Mitigation such as the course of action. IODEF contains an optional attribute called restriction. The purpose of this attribute is to inform the receiver how they should deal with this information. The suggested values are public, need-to-know and private. However, using this property will not enforce the receiver to apply it. We found that IODEF incident class properties from that constitute incident are similar to the properties already existing in STIX 1.2. It has a "Description" property that is a free-text filed to describe the incident, a "Method" property which describes the techniques used to conduct the attack and the existing weakness, contact information and other properties which match in general our understanding of what kinds of information exists and the resulting consequences of disclosing this information. Finally, our analysis pointed out that the risks of sharing cyber incident information are still the same even when using different standards of representing cyber threat information.

3.4 Conclusion

In this chapter, 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. The following chapter will extend this work by proposing a new risk assessment model for sharing cyber threat information validated by empirical evaluation.



Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
Version	×	×	×	×	×	This refers to the report not the incident.	N/A	0	0
URL	×	×	×	×	×	This refers to the report not the incident.	N/A	0	0
Title	×	✓	✓	×	×	This is a free text field which can refer to particular business information.	T5, T1	2+2	2+2
External_ID	×	✓	×	×	×	This refers to the report not the incident.	N/A	0	0
Time	✓	×	✓	QI	×	The "TimeType" class has sensitive and QI properties.	T1, T5, T3, T4, T22, T24, T26	0	2+2
Description	✓	✓	✓	✓	*	This is a free text field which can refer to particular business information and may contain sensitive and identifying information.	T5, T1, T4, T19, T20, T21, T24, T25	2+2	2+2
Short_Description	✓	✓	✓	✓	*	This is a free text field which can refer to particular business information and may contain sensitive and identifying information.	T5, T1, T4, T19, T20, T21, T24, T25	2+2	2+2
Categories	✓	✓	×	×	×	This property is high level category which is part of the report not the incident.	N/A	0	0
Reporter	✓	✓	✓	✓	✓	The identity of the reporter can be revealed.	T1, T7, T8, T10, T26	4+2	2+2
Responder	✓	✓	✓	✓	✓	The identity of the responder can be revealed.	T7, T1, T8, T10, T26	4+2	2+2
Coordinator	✓	✓	✓	✓	✓	The identity of the coordinator can be revealed.	T7, T1, T8, T10, T26	4+2	2+2
Victim	✓	✓	✓	✓	✓	Can refer to a particular critical infrastructure and the identity of the victim can be revealed.	T1, T5, T7, T8, T9, T11, T12, T21, T24, T25, T26	4+2	2+3
Affected_Assets	✓	✓	✓	×	×	The "AffectedAssetsType" class has sensitive properties.	T1, T3, T4, T6, T9, T10, T11, T19, T20, T21, T22, T23, T24, T25, T26	0	2+2
Impact_Assessment	✓	✓	✓	×	×	The "ImpactAssessmentType" class has sensitive properties.	T1, T3, T4, T12, T13, T14,	0	2+3
Status	✓	✓	✓	×	×	This field can refer to particular business information.	T3, T5, T9, T12, T22, T25, T26	0	2+2
Related_Indicators	✓	✓	✓	✓	✓	This can refer to specific information about the incident and adversary Tactics, Techniques, and Procedures (TTPs). And may contain identifying information about the adversary.	T1, T3, T5, T8, T9, T14, T22, T26	2+2	2+2
Related_Observables	✓	✓	✓	✓	✓	The "RelatedObservablesType" class has sensitive and identifying properties.	T1, T3, T5, T22, T26	2+2	2+2

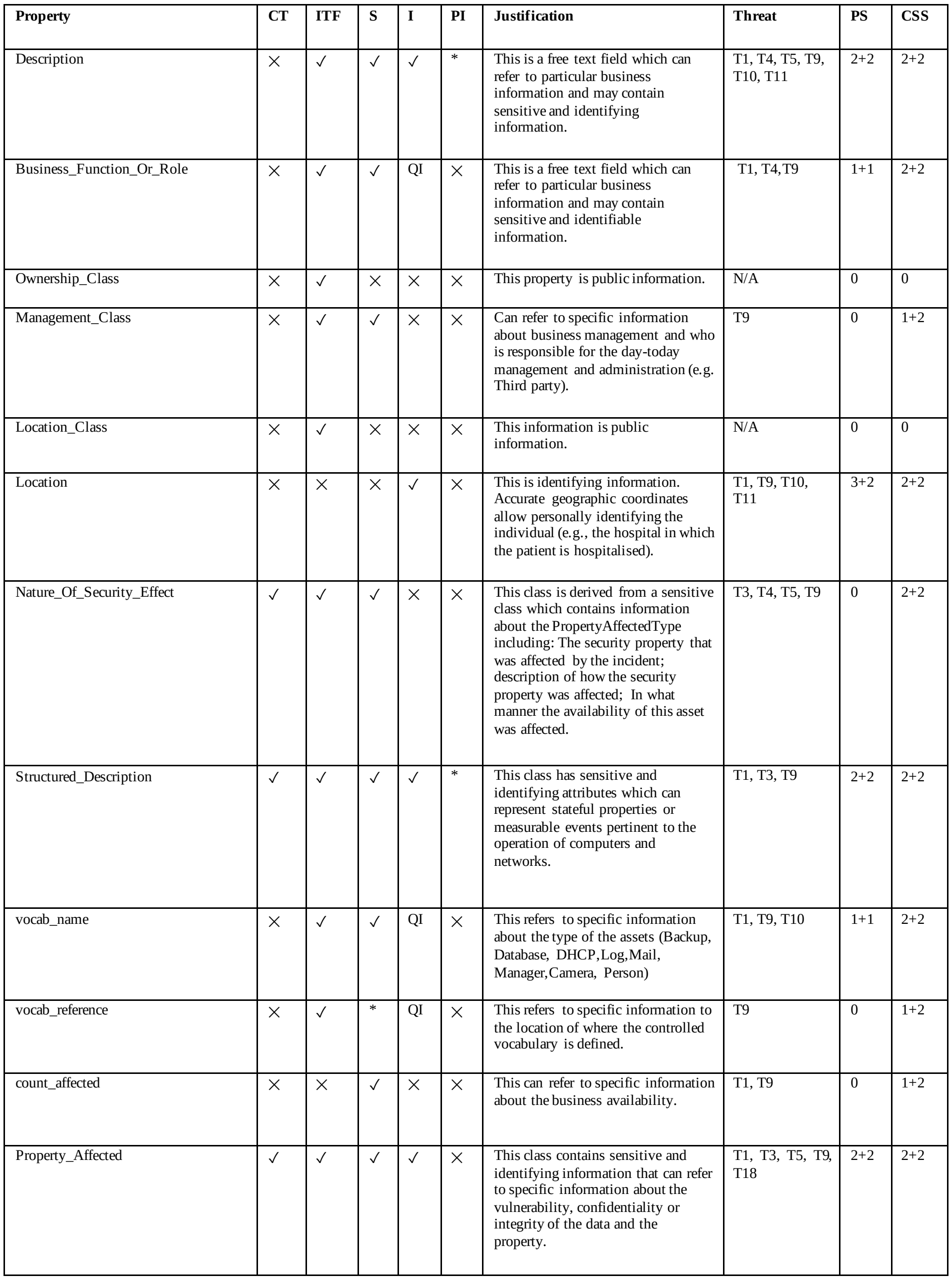
Table 17: Full analysis for Class “IncidentType”

Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
Leveraged_TTPs	✓	✓	✓	✓	✓	This can refer to the adversary Tactics, Techniques, and Procedures (TTPs) and victim weaknesses and it specifies the identity of the information source.	T1, T5, T9, T23, T24	2+2	2+2
Attributed_Threat_Actors	✓	✓	✓	✓	✓	The "AttributedThreatActorsType" class has sensitive and identifyale properties.	T1, T5, T9, T24	2+2	3+3
Intended_Effect	✓	✓	✓	×	✓	Refer to the goal of the attack.	T1, T9	2+2	2+2
Security_Compromise	×	×	✓	×	×	Identifies whether critical information was leaked, and it can be sensitive.	T5, T9, T12, T22, T26	0	3+3
Discovery_Method	×	×	*	×	×	Not generally but there is revealing of some security controls.	T5, T12, T24	0	2+2
Related_Incidents	✓	✓	✓	✓	✓	The "RelatedIncidentsType" class has sensitive and identifier properties	T1, T5, T9, T12, T22, T24, T25, T26	4+1	2+2
COA_Requested	✓	✓	✓	✓	✓	This can refer to specific information about the business, and how the organisation can return to business as usual. Recovery operation includes their own security risks used in the targeted attack.	T9, T16, T1, T17, T5, T9, T19, T24	2+2	2+2
COA_Taken	✓	✓	✓	✓	✓	This can refer to specific information about the business, and how the organisation can return to business as usual. Recovery operation includes their own security risks used in the targeted attack.	T16, T1, T17, T5, T9, T19, T24	2+2	2+2
Confidence	✓	✓	×	✓	✓	The "ConfidenceType" class has Source field which specifies the source of this confidence assertion.	T12	2+2	2+2
Contact	✓	✓	×	✓	✓	The identity of the point of contact and personnel involved in the incident can be revealed.	T1, T7, T9	4+2	2+2
History	✓	×	✓	✓	✓	The "History" class has sensitive and identifible properties.	T1, T5, T9	2+2	2+2
Information_Source	✓	✓	✓	✓	✓	This can refer to specific information about the business policy and the tools utilised for this instance. Can reveal the identity of the information source and the individual contributing sources involved in this instance.	T1, T5	4+2	2+2
Handling	✓	✓	×	✓	✓	This class contains fields such as "Information_Source" field which gives details about the source of this entry.	T1, T5	2+2	2+2
Related_Packages	✓	×	×	✓	✓	This refers to the report not the incident. The "Information_Source" field specifies the source of the	T1, T5	4+1	2+2

Table 17: Full analysis for Class “IncidentType” (cont.)

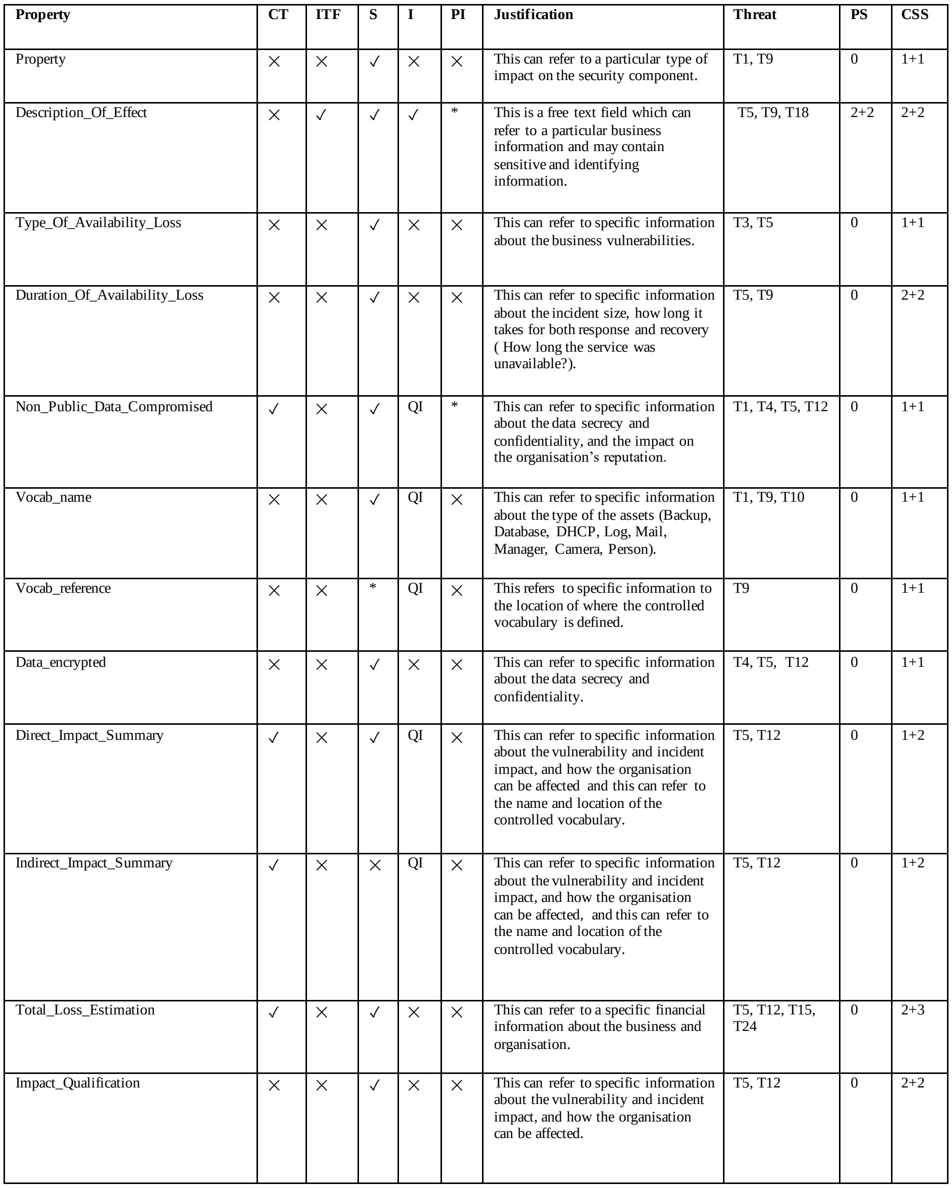
Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
						information about the relationship between the two components.			
First_Malicious_Action	×	×	*	×	×	Not generally but series of this information may lead to patterns of behavior and location of the attacker.	N/A	0	0
Initial_Compromise	×	×	×	×	×	This can refer to specific information about the business and potential of compromise.	N/A	0	2+2
First_Data_Exfiltration	×	×	×	QI	×	This can be combined with other sources (Media) to create unique identifier.	T5	0	2+2
Incident_Discovery	×	×	✓	×	×	This can refer to specific information about the business, when the company started to solve the problem and when the organisation released statement warning customers if needed.	T5	0	2+2
Incident_Opened	×	×	✓	×	×	This can refer to specific information about the current security and incident response policy.	T5	0	2+2
Containment_Achieved	×	×	✓	×	×	This can refer to specific information about the incident size, how long it takes for both response and recovery (The delayed containment strategy is dangerous because an attacker could escalate unauthorised access or compromise other systems).	T3, T5	0	2+2
Restoration_Achieved	×	×	✓	×	×	This can refer to specific information about the incident size, how long it takes for recovery.	T5	0	2+2
Incident_Reported	×	×	✓	×	×	This can refer to specific information about the organisation and when the incident happened. This information is sensitive as it confirms the possibility of leaking or impacting the business.	T4, T5	0	2+2
Incident_Closed	×	×	✓	×	×	This can refer to specific information about the incident size, how long it takes for both response and recovery.	T5	0	2+2
Affected_Asset	✓	✓	✓	✓	×	This can refer to specific information about the organisation and the business (knowing the appropriate actions are taken for the affected asset). This class has sensitive and identifying information.	T1, T3, T4, T5, T6, T9, T10, T11, T12	4+2	2+2
Type	✓	×	✓	QI	×	This can refer to a specific business type such as critical IT infrastructure in key sector of the economy. Quasi-identifier: it can be combined with other property to create a unique identifier.	T1, T9, T10	1+1	1+2

Table 17: Full analysis for Class “IncidentType” (cont.)



Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
Description	✕	✓	✓	✓	*	This is a free text field which can refer to particular business information and may contain sensitive and identifying information.	T1, T4, T5, T9, T10, T11	2+2	2+2
Business_Function_Or_Role	✕	✓	✓	QI	✕	This is a free text field which can refer to particular business information and may contain sensitive and identifiable information.	T1, T4,T9	1+1	2+2
Ownership_Class	✕	✓	✕	✕	✕	This property is public information.	N/A	0	0
Management_Class	✕	✓	✓	✕	✕	Can refer to specific information about business management and who is responsible for the day-today management and administration (e.g. Third party).	T9	0	1+2
Location_Class	✕	✓	✕	✕	✕	This information is public information.	N/A	0	0
Location	✕	✕	✕	✓	✕	This is identifying information. Accurate geographic coordinates allow personally identifying the individual (e.g., the hospital in which the patient is hospitalised).	T1, T9, T10, T11	3+2	2+2
Nature_Of_Security_Effect	✓	✓	✓	✕	✕	This class is derived from a sensitive class which contains information about the PropertyAffectedType including: The security property that was affected by the incident; description of how the security property was affected; In what manner the availability of this asset was affected.	T3, T4, T5, T9	0	2+2
Structured_Description	✓	✓	✓	✓	*	This class has sensitive and identifying attributes which can represent stateful properties or measurable events pertinent to the operation of computers and networks.	T1, T3, T9	2+2	2+2
vocab_name	✕	✓	✓	QI	✕	This refers to specific information about the type of the assets (Backup, Database, DHCP,Log,Mail, Manager,Camera, Person)	T1, T9, T10	1+1	2+2
vocab_reference	✕	✓	*	QI	✕	This refers to specific information to the location of where the controlled vocabulary is defined.	T9	0	1+2
count_affected	✕	✕	✓	✕	✕	This can refer to specific information about the business availability.	T1, T9	0	1+2
Property_Affected	✓	✓	✓	✓	✕	This class contains sensitive and identifying information that can refer to specific information about the vulnerability, confidentiality or integrity of the data and the property.	T1, T3, T5, T9, T18	2+2	2+2

Table 17: Full analysis for Class “IncidentType” (cont.)

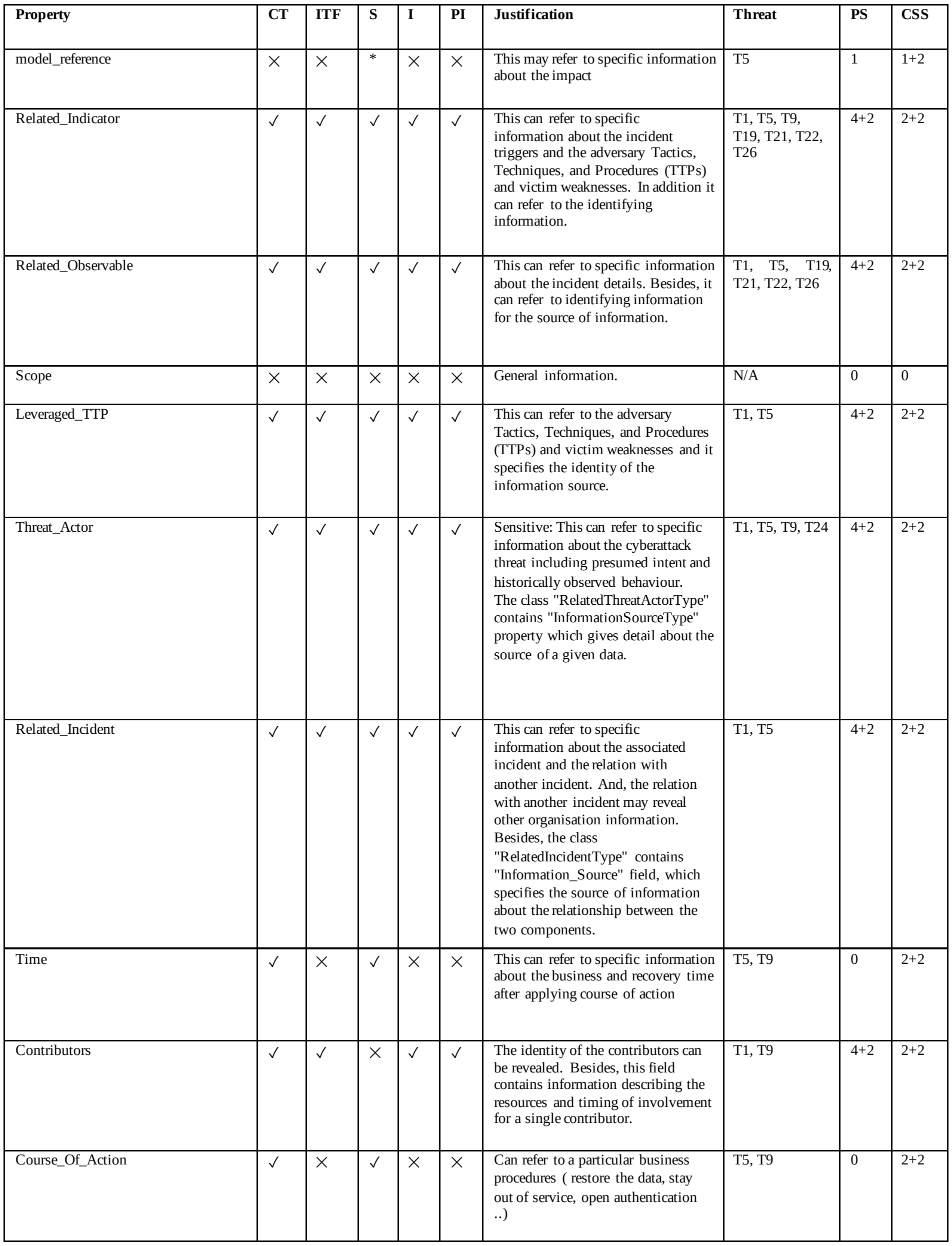


Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
Property	×	×	✓	×	×	This can refer to a particular type of impact on the security component.	T1, T9	0	1+1
Description_Of_Effect	×	✓	✓	✓	*	This is a free text field which can refer to a particular business information and may contain sensitive and identifying information.	T5, T9, T18	2+2	2+2
Type_Of_Availability_Loss	×	×	✓	×	×	This can refer to specific information about the business vulnerabilities.	T3, T5	0	1+1
Duration_Of_Availability_Loss	×	×	✓	×	×	This can refer to specific information about the incident size, how long it takes for both response and recovery (How long the service was unavailable?).	T5, T9	0	2+2
Non_Public_Data_Compromised	✓	×	✓	QI	*	This can refer to specific information about the data secrecy and confidentiality, and the impact on the organisation's reputation.	T1, T4, T5, T12	0	1+1
Vocab_name	×	×	✓	QI	×	This can refer to specific information about the type of the assets (Backup, Database, DHCP, Log, Mail, Manager, Camera, Person).	T1, T9, T10	0	1+1
Vocab_reference	×	×	*	QI	×	This refers to specific information to the location of where the controlled vocabulary is defined.	T9	0	1+1
Data_encrypted	×	×	✓	×	×	This can refer to specific information about the data secrecy and confidentiality.	T4, T5, T12	0	1+1
Direct_Impact_Summary	✓	×	✓	QI	×	This can refer to specific information about the vulnerability and incident impact, and how the organisation can be affected and this can refer to the name and location of the controlled vocabulary.	T5, T12	0	1+2
Indirect_Impact_Summary	✓	×	×	QI	×	This can refer to specific information about the vulnerability and incident impact, and how the organisation can be affected, and this can refer to the name and location of the controlled vocabulary.	T5, T12	0	1+2
Total_Loss_Estimation	✓	×	✓	×	×	This can refer to a specific financial information about the business and organisation.	T5, T12, T15, T24	0	2+3
Impact_Qualification	×	×	✓	×	×	This can refer to specific information about the vulnerability and incident impact, and how the organisation can be affected.	T5, T12	0	2+2

Table 17: Full analysis for Class "IncidentType" (cont.)

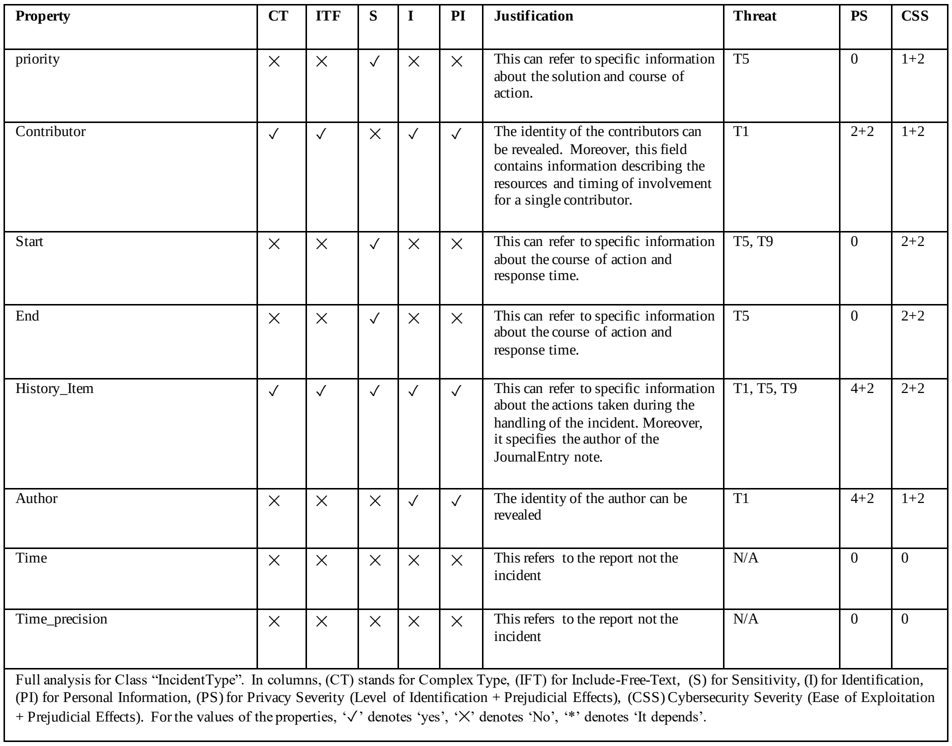
Property	**CT**	**ITF**	**S**	**I**	**PI**	**Justification**	**Threat**	**PS**	**CSS**
Effects	✓	×	✓	×	×	This can refer to specific information about the vulnerability and incident impact, and how the organisation can be affected. e.g. Customer breach notifications, Regulatory compliance (fines), damage business' reputation	T5, T12, T24	0	2+2
External_Impact_Assessment_Model	✓	×	✓	×	×	This can refer to specific information about the vulnerability and incident impact, and how the organisation can be affected. and quantify the impact of cyber-attacks (Reliability).	T3, T12	0	2+2
Asset_Losses	×	×	✓	×	×	Can refer to specific information about the business and how the organisation has been affected.	T5	0	2+2
Business-Mission_Disruption	×	×	✓	×	×	Can refer to specific information about the business and how the organisation has been affected.	T3, T5, T12	0	2+2
Response_And_Recovery_Costs	×	×	✓	×	×	This can refer to specific financial information about the business and the incident recovery cost.	T5	0	2+2
Loss_Of_Competitive_Advantage	×	×	✓	×	×	This can refer to specific information about the incident damage.	T5	0	2+2
Brand_And_Market_Damage	×	×	✓	QI	×	This can refer to specific information about the business status after the incident and can be combined with other properties to identify the organisation.	T5, T12	0	2+2
Increased_Operating_Costs	×	×	✓	×	×	This can refer to specific financial information about the business and the incident recovery cost.	T5	0	2+2
Legal_And_Regulatory_Costs	×	×	✓	×	×	This can refer to specific financial information about the business and the incident recovery cost.	T5, T12	0	2+2
Initial_Reported_Total_Loss_Estimation	✓	×	✓	×	×	This can refer to estimated financial loss for the incident.	T5, T12	0	2+2
Actual_Total_Loss_Estimation	✓	×	✓	×	×	This can refer to estimated financial loss for the incident.	T5, T12, T24	0	2+2
Amount	×	×	*	×	×	This can refer to specific financial information when combine it with iso_currency_code property.	T5, T12	0	1+2
iso_currency_code	×	×	*	×	×	This can refer to specific financial information when combine it with amount property.	T5	0	1+2
Effect	×	×	✓	×	×	This can refer to specific information about the business status and the attack effect	T5, T12, T24	0	1+2
Model_name	×	×	*	×	×	This may refer to specific information about the impact	T5	0	1+2

Table 17: Full analysis for Class “IncidentType” (cont.)



Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
model_reference	×	×	*	×	×	This may refer to specific information about the impact	T5	1	1+2
Related_Indicator	✓	✓	✓	✓	✓	This can refer to specific information about the incident triggers and the adversary Tactics, Techniques, and Procedures (TTPs) and victim weaknesses. In addition it can refer to the identifying information.	T1, T5, T9, T19, T21, T22, T26	4+2	2+2
Related_Observable	✓	✓	✓	✓	✓	This can refer to specific information about the incident details. Besides, it can refer to identifying information for the source of information.	T1, T5, T19, T21, T22, T26	4+2	2+2
Scope	×	×	×	×	×	General information.	N/A	0	0
Leveraged_TTP	✓	✓	✓	✓	✓	This can refer to the adversary Tactics, Techniques, and Procedures (TTPs) and victim weaknesses and it specifies the identity of the information source.	T1, T5	4+2	2+2
Threat_Actor	✓	✓	✓	✓	✓	Sensitive: This can refer to specific information about the cyberattack threat including presumed intent and historically observed behaviour. The class "RelatedThreatActorType" contains "InformationSourceType" property which gives detail about the source of a given data.	T1, T5, T9, T24	4+2	2+2
Related_Incident	✓	✓	✓	✓	✓	This can refer to specific information about the associated incident and the relation with another incident. And, the relation with another incident may reveal other organisation information. Besides, the class "RelatedIncidentType" contains "Information_Source" field, which specifies the source of information about the relationship between the two components.	T1, T5	4+2	2+2
Time	✓	×	✓	×	×	This can refer to specific information about the business and recovery time after applying course of action	T5, T9	0	2+2
Contributors	✓	✓	×	✓	✓	The identity of the contributors can be revealed. Besides, this field contains information describing the resources and timing of involvement for a single contributor.	T1, T9	4+2	2+2
Course_Of_Action	✓	×	✓	×	×	Can refer to a particular business procedures (restore the data, stay out of service, open authentication ..)	T5, T9	0	2+2

Table 17: Full analysis for Class “IncidentType” (cont.)



Property	CT	ITF	S	I	PI	Justification	Threat	PS	CSS
priority	✕	✕	✓	✕	✕	This can refer to specific information about the solution and course of action.	T5	0	1+2
Contributor	✓	✓	✕	✓	✓	The identity of the contributors can be revealed. Moreover, this field contains information describing the resources and timing of involvement for a single contributor.	T1	2+2	1+2
Start	✕	✕	✓	✕	✕	This can refer to specific information about the course of action and response time.	T5, T9	0	2+2
End	✕	✕	✓	✕	✕	This can refer to specific information about the course of action and response time.	T5	0	2+2
History_Item	✓	✓	✓	✓	✓	This can refer to specific information about the actions taken during the handling of the incident. Moreover, it specifies the author of the JournalEntry note.	T1, T5, T9	4+2	2+2
Author	✕	✕	✕	✓	✓	The identity of the author can be revealed	T1	4+2	1+2
Time	✕	✕	✕	✕	✕	This refers to the report not the incident	N/A	0	0
Time_precision	✕	✕	✕	✕	✕	This refers to the report not the incident	N/A	0	0

Full analysis 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 (Level of Identification + Prejudicial Effects), (CSS) Cybersecurity Severity (Ease of Exploitation + Prejudicial Effects). For the values of the properties, ‘✓’ denotes ‘yes’, ‘✕’ denotes ‘No’, ‘*’ denotes ‘It depends’.

Table 17: Full analysis for Class “IncidentType” (cont.)

Chapter 4

Risk Assessment of Sharing Cyber Threat Intelliegnce

In this chapter[1], we present a quantitative risk model to assess the risk of sharing CTI datasets enabled by sharing with different entities in various situations. The model enables the identification of the threats and evaluation of the impacts of disclosing this information. We present three use cases that help to determine the risk level of sharing a CTI dataset and consequently, the mitigation techniques to enable responsible sharing. Risk identification and evaluation have been validated using experts' opinions.

4.1 Introduction

In the previous chapter, we performed a comprehensive analysis of incident reporting information through the STIX incident model to identify the threats of disclosing sensitive and identifying information. We identified the threats associated with each property and evaluated those for severity in both the privacy and cybersecurity dimension. The next step is to provide a risk model for evaluating

[1] This chapter is based on the conference paper "Risks of Sharing Cyber Incident Information" [207]

the risk of sharing CTI datasets.

In this chapter, we will propose a specific quantitative risk model for evaluating the risk of sharing CTI datasets. This model builds on the identification and partial assessment of threats in cyber incident information sharing in Chapter 3. This model will help improve the decision making on sharing CTI information with multiple entities. During the evaluation phases, we take into consideration the threats of sharing each attribute in the CTI dataset and the likelihood of such threats occurring and the level of trust in the receiving party. Sharing CTI datasets has specific consequences which make organisations reluctant to share, as discussed in detail in Section 3.2.4. The barriers can be: (1) the probability of undesirable information disclosure increases when shared with organisations that do not have a high level of trust or when sharing with the public, (2) CTI datasets can contain various kinds of information such as personal, organisational, financial, business, and cybersecurity information [180]. Thus, evaluating the risk of sharing CTI datasets containing critical information such as the existing vulnerabilities is a challenge, especially with the evolving cyber threat landscape and sophisticated cyber-attacks for various business sectors. When considering the different sources of CTI information and the intention to share with various entities, a risk assessment model is needed. By evaluating the associated risk of sharing CTI datasets, organisations would know how critical their CTI datasets are before sharing [180] and use the right methods and processes to manage the risk to respect the organisation's acceptable risk level. In addition, they need to obtain legal compliance as the General Data Protection Regulation (GDPR) [149] mandates organisations to undertake risk assessments and fulfil security mitigation controls – this is discussed in detail in Chapter 5.

The remainder of this chapter is organised as follows. Sections 4.2 to 4.4 describe the steps of the methodology to build the model. Section 4.5 gives several use cases of sharing CTI datasets to validate the model by involving cybersecurity experts. Section 4.6 describes threats to validity of the model. Finally, Section 4.7 summarises this chapter.

4.2 Associated Risk Model (ARM)

In this section, we present our associated risk model (ARM). The first step in our ARM procedure is to examine the dataset. In this step, we will be indicating the roles that the various attributes may play: they could contain sensitive information, or help to identify people and organisations. We then point out threats, using the ENISA threat taxonomy as described in Section 3.1.2. We compute the severity for each property in the dataset because if there is a disclosure of sensitive and critical information, there would be a risk that an associated threat could exploit the system, and the organisation may face an unexpected cybersecurity attack, reputational damage and legal consequences.

We have precisely identified the associated threats by analysing each property in the STIX 1.2 incident model separately and mapping it to the ENISA threat taxonomy as described in Chapter 3. Then, for each property we have calculated the severity value that was assigned in Chapter 3. After identifying the potential threats, we can derive the level of associated risk for this sharing by estimating the likelihood of the threats in case of property disclosure.

Our goal is to reduce the risk value by selecting the appropriate privacy preserving techniques to improve the sharing between organisations. Figure 8 illustrates the flow chart of ARM which describes the risk assessment steps, including identification of risks through the disclosure of the shared dataset properties, and their total risk value through the analysis of threats mapped based on the disclosed properties.

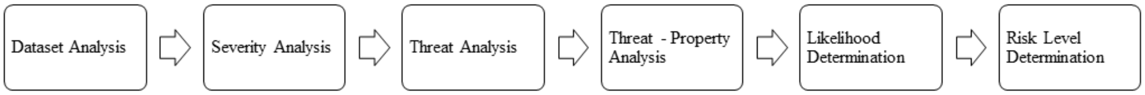


Figure 8: ARM Steps

4.2.1 Dataset analysis

First, we need to identify the associated risk of disclosing any property of the shared CTI dataset. Each property may have a different severity level in an organisation. In Chapter 3, we have estimated the cybersecurity severity score for each property in the STIX 1.2 incident model [208]. The severity score range is [1,8], where 1 is the lowest level of severity and 8 is the highest level of severity. Based on the severity score, severity was assigned to four impact levels: negligible, limited, significant and maximum which can be represented as 10, 50, 75 and 100. There are limitations to using ordinal scales in risk assessments [209]. In [209], researchers discuss the possibility of bias and subjectivity coming from different levels of professionals' experiences and inconsistency in understanding each incident's factors and indicators. However, various risk assessment standards considered "best practice" to estimate cyber risks use ordinal scores. For example, the NIST 800-30 standard for conducting information systems risk assessments [157] is based on ordinal scores. Also, the main advantage of using the ordinal scale is the ease of comparison between variables. In this model, estimating the severity level was defined based on CNIL methodology. CNIL methodology [182] has a detailed and precise definition for each selected score. We have extended this method to the cybersecurity risks, estimating cybersecurity severity (CSS) score. However, changing the scale would not affect the model as it also depends on the organisation risk profile, so we will leave the organisation to decide how to handle the risks and the acceptance level. Also, changing the numbers of ranks on the severity scale would not change the final evaluation process as it would be related to the organisation risk profiles.Let each property be represented as a single bit in the property vector:

$$\overrightarrow{P} = \{P_i\} \in \{0,1\} \ \forall i \ , \ i = 1, \ 2, \ \ldots n \tag{4}$$

Here, P_i represents an individual property. The value 1 indicates the existence of this property in the shared dataset, otherwise it is 0. Because disclosing any property in the shared dataset is a potential risk, we include all properties into our analysis. If we are fully sharing a dataset with 10 properties, we set n to 10 and $P_i = 1 \ \forall n$.

4.2.2 Threat analysis

The second step in our model is to perform a threat analysis, which consists of identifying the potential threat action that may exploit the system or the organisation based on the CTI information disclosure. Information about threats can be collected from the organisation's CTI database and threat taxonomies (see Section 2.4.5) which define a list of potential threats to the organisation.
Let each threat be represented as a single bit in the threat vector:

$$\overrightarrow{T} = \{T_j\} \in \{0,1\} \forall j \ , \ j = 1, \ 2, \ \ldots m. \tag{5}$$

Here T_j represents an individual threat, the value 1 indicates the presence of this threat when sharing the CTI dataset and otherwise it is 0. Thereafter, based on the CTI dataset disclosure and the associated threats, we can match threats to the CTI property and estimate the likelihood of a threat occurring based on disclosure of CTI information. For adversarial threats, there are several types of adversarial attempt to exploit the organisation's systems and infrastructures. These could be individual (outsider, insider), group (Ad hoc and Established), organisation (competitor, supplier, partner and customer) and nation-state [210]. When the adversary has the ability, has sufficient resources, and can create opportunities to perform multiple successful and synchronised attacks, the likelihood of performing an attack will be very high, or we can consider it 1. In this thesis, we assume that the adversary is assertive and powerful. Therefore, the likelihood values L_{ij} are based on how easy it is for a threat to be executed by a motivated and powerful adversary. This likelihood can adopt three values: low, medium, high represented as 0.1, 0.5 and 1. In case there is no risk, we assign value L_{ij}=0. In the previous step, there will be a subjective factor - expert judgment - because of the diverse perception of associated threats for each property, what impact that would have on the organisation and the likelihood of an event happening. The judgment of the likelihood value would be based on the available context which might be related to the organisational aspects (e.g., business sector, geographic location), perpetrator motivation to cause a cybersecurity attack combined with their resources and abilities. In critical infrastructures, it is reasonable to assume that a motivated perpetrator exists this why we need to focus on the likelihood of successfully exploiting existing vulnerabilities. Each CTI dataset comes from a different business sector, context and countries that could create different associated threats such as the legal assessment. Therefore, a specific way of calculating the associated risk and defining each risk level in terms of expected impact and expected techniques to share securely might be a mandatory pre-requirement for sharing CTI datasets. For example, the impact of gaining access over the ATM control system in order to withdraw money is different than the impact of gaining control over CCTV cameras in critical infrastructure.

4.2.3 Total Associated Risk (TAR)

Total Associated Risk (TAR) is the sum of associated subrisks of disclosing CTI information and can be computed as follows:

$$TAR = \sum_{i=1}^{n}\sum_{j=1}^{m} L_{ij} * S_i * \ P_j * \ T_i \ whereTAR \in \ \mathbb{R}^{+} \qquad (6)$$

where n represents the number of the properties, m represents the number of the threats, L_{ij} represents the likelihood of the presence of the threat i when disclosing the property j and S_i represents the severity score. The likelihood values L_{ij} represent how easy it is for a powerful and motivated adversary to execute threat j knowing property i.

Once TAR has been computed, the organisation becomes aware of how this could provide the appropriate information to decision makers about how to make a clear decision about sharing this dataset and how to evaluate the associated risk.

4.3 Evaluation

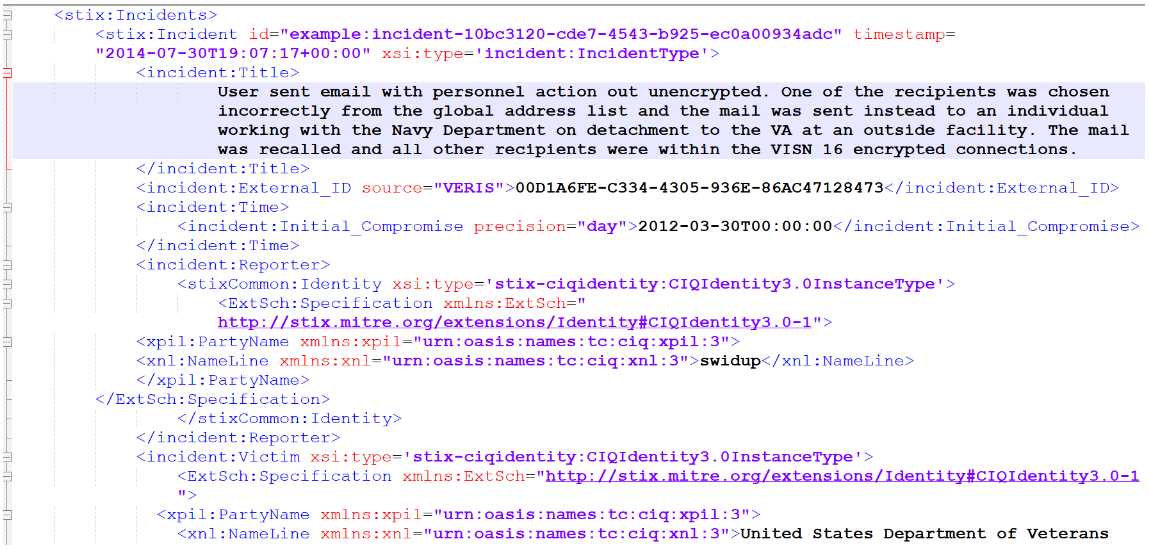
To evaluate the ARM model, we have conducted an experiment on a repository containing STIX documents [122] and another experiment on three case studies that were analysed manually using our model by independent experts.

4.3.1 Experiment set up

Dataset information

CTI datasets are collected using different platforms that are either open source or commercially based. In this chapter, we have used the largest known public dataset of STIX incidents as provided by MITRE [211] [122]. A sample of a STIX incident report is shown in Figure 9.

The dataset consists of 4788 STIX incident reports. Our experiment focuses on evaluating the risk of sharing these incidents. We have implemented a parser to parse each file in the dataset to extract the properties included in the report and look at the value of each property in order to evaluate them against the associated risk model.



```
<stix:Incidents>
    <stix:Incident id="example:incident-10bc3120-cde7-4543-b925-ec0a00934adc" timestamp=
    "2014-07-30T19:07:17+00:00" xsi:type='incident:IncidentType'>
        <incident:Title>
                User sent email with personnel action out unencrypted. One of the recipients was chosen
                incorrectly from the global address list and the mail was sent instead to an individual
                working with the Navy Department on detachment to the VA at an outside facility. The mail
                was recalled and all other recipients were within the VISN 16 encrypted connections.
        </incident:Title>
        <incident:External_ID source="VERIS">00D1A6FE-C334-4305-936E-86AC47128473</incident:External_ID>
        <incident:Time>
            <incident:Initial_Compromise precision="day">2012-03-30T00:00:00</incident:Initial_Compromise>
        </incident:Time>
        <incident:Reporter>
            <stixCommon:Identity xsi:type='stix-ciqidentity:CIQIdentity3.0InstanceType'>
                <ExtSch:Specification xmlns:ExtSch="
                http://stix.mitre.org/extensions/Identity#CIQIdentity3.0-1">
        <xpil:PartyName xmlns:xpil="urn:oasis:names:tc:ciq:xpil:3">
        <xnl:NameLine xmlns:xnl="urn:oasis:names:tc:ciq:xnl:3">swidup</xnl:NameLine>
        </xpil:PartyName>
    </ExtSch:Specification>
            </stixCommon:Identity>
        </incident:Reporter>
        <incident:Victim xsi:type='stix-ciqidentity:CIQIdentity3.0InstanceType'>
            <ExtSch:Specification xmlns:ExtSch="http://stix.mitre.org/extensions/Identity#CIQIdentity3.0-1
            ">
          <xpil:PartyName xmlns:xpil="urn:oasis:names:tc:ciq:xpil:3">
            <xnl:NameLine xmlns:xnl="urn:oasis:names:tc:ciq:xnl:3">United States Department of Veterans
```

Figure 9: Sample of STIX Incident Report

Parameter Settings

This automated analysis is entirely based on the presence or absence of specific properties, as most are optional in a STIX 1.2 incident report. We only analyse the content to see whether the property is essentially absent. The process starts with parsing the STIX incident reports and extracting the properties with their associated values. Some properties have an empty or 'Unknown' value. Those properties have been excluded from the evaluation. By doing so, we are providing more accurate results by removing the non-useful information from the evaluation. For this dataset, we were able to perform a simple analysis to capture context and sensitivity. However, for a more complex and more extensive dataset, a text mining technique enhances the analysis of properties such as "incident description". This technique will help to extract information for categorising the nontrivial patterns and classify the sensitivity of this information [212]. Table 18 shows the property list that we have extracted from the parsed files. Also, we need to define the list of threats associated with the risk of sharing the CTI dataset.

Based on Chapter 3, we defined the list of threats associated with disclosing these properties. Table 19 shows the list of threats.

Property List		
Title	Coordinator	Related Observables
Time	Victim	Leveraged TTPs
Description	Affected Assets	Attributed Threat Actors
Short Description	Impact Assessment	Intended Effect
Reporter	Status	Security Compromise
Responder	Related Indicators	Discovery Method
Related Incidents		

Table 18: Property List

Threat List	
Social engineering	Loss of reputation
Failure to meet contractual requirements	Manipulation of information
Violation of laws or regulations	Misuse of information/systems
Compromising confidential information	Failed business process
Identity theft (Identity Fraud/ Account)	Man in the middle/ Session hijacking
Unauthorised activities	Generation and use of rogue certificates
Targeted attacks (APTs etc.)	Abuse of authorisations
Judiciary decisions/court orders	

Table 19: Threat list

Results

One of our goals is to evaluate the cybersecurity risk of sharing CTI data. To assess the effectiveness of achieving this goal, we presents some measurements based on the results obtained from the associated risk values of the sample dataset. These measurement are:

- The maximum possible total associated risk is 1320.
- The lowest associated risk value in this dataset is 70.
- The maximum associated risk value in this dataset is 345.
- The average associated risk is 291.

For the maximum value, we have used the worst-case scenario. We assumed that we have an incident report which contains all the properties in this dataset along with their values. The histogram of the data processed is presented in Figure 10. The histogram has one peak for the risk range between 281 and 300. The minimum value is 70 and the maximum value is 345. There are no gaps or extreme outliers.

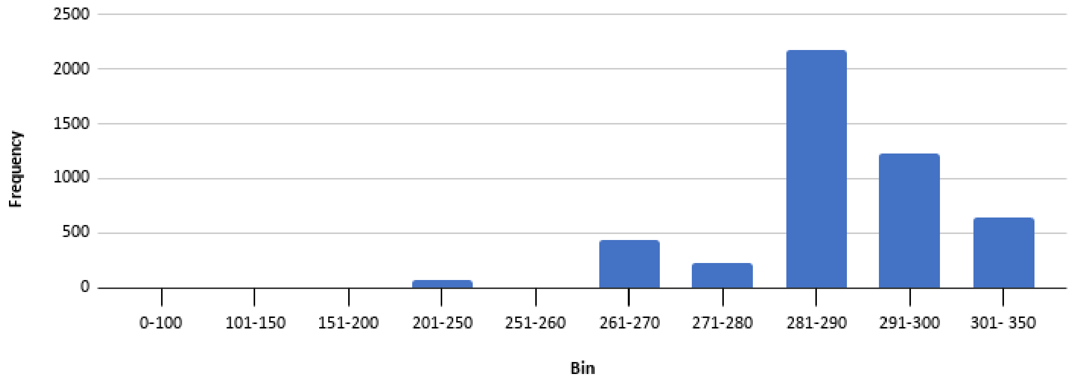


Figure 10: Histogram analysis of the sample dataset

The analysis as a whole is also essentially a worst case one: it assumes any information contained in a property represents all the associated threats. We can observe that when the risk value is low then the incident report is unlikely to be useful when sharing for analysis. For example, the incident with title “advertising servers were compromised and made to serve up malware (darkleech)” has a risk value of 115. This report only contains general and public information without specific information about the victim, technical or business details.

On the other hand, with a high-risk value the report will contain more information about the incident, which may be useful for analysis. For example, the incident report with the title “Embedding Scripts in Non-Script Elements” contains technical information about the attack pattern which has a high severity impact, the victim information and location, in addition to the affected assets information and their properties such as the loss of availability for days.

4.4 Expert selection

In this study, we have developed three use cases aiming to validate our model. Two use cases were analysed by independent experts with different levels of experience working on cybersecurity and privacy during a privacy workshop. Also, we have asked PhD students (third year) during a PhD summer school to fill a questionnaire, where all PhD students are working in cybersecurity. The third use case was analysed by professionals with several years of experience in Red Team activity, penetration testing and industrial control systems.

4.5 Case Studies

The presented ARM is here tested through three use case studies. Case study 1 discusses sharing a CTI dataset for correlation purposes, while case study 2 discusses sharing a CTI dataset for aggregation purposes and case study 3 discusses sharing for detection purposes. In all case studies, we consider sharing with trusted and untrusted entities.

4.5.1 Use Case 1: CTI contains malware information & personal information - sharing for detections

This scenario consists of two cyber threat companies, CyberA and CyberB. CyberA has been attacked by specific malware. This malware was designed to steal encrypted files - and was even able to recover files that had been deleted. CyberA wants to share this incident dataset with others in their sharing community. The purpose of this sharing is to let recipients check if they have the same malware on their system.

Table 20 shows the sample CTI dataset, which contains the properties that might be shared.

Property	**Value**
TTP Malware Type	Capture Stored Data, Remote Access Trojan
Indicator Name	File hash for malicious malware
Indicator Description	This file hash indicates that a sample of malware alpha is present.
Indicator Value	Hashes.’SHA-256’= ’ef537f25c895bfa7jfdhfjns73748hdfjkk5d89fjfer8fjkdndkjn7yfb6c’ Windows-registry-key:= “HKEY_LOCAL_MACHINE\\SYSTEM\\ControlSet001\\Services\\MSADL3”
Vulnerability	CVE-2009-3129, CVE-2008-4250, CVE-2010-3333, CVE-2012-0158, CVE-2011-3544
Incident Title	Incident associated with CyberA campaign. The malware was designed to steal encrypted files - and was even able to recover files that had been deleted.
Date	2012-01-01T00:00:00
Reporter Name	Alex John
Reporter Email Address	alex@pro-it.com
Reporter Address	US-LA
Victim Name	CyberA / The CEO Device
Victim sector	Financial sector
Victim Device	IP address: 146.227.239.19
Victim Email Address	cybera@cyber.com / ceo-cybera@cyber.com
Victim Address	CyberA Ltd, IT Department, LONDON, W5 5YZ
Affected Assets Type	Desktop, Mobile phone, Router, Server, Person
Affected Assets Property	Confidentiality (Classified, Internal, Credentials, Secrets, System) Integrity (Software installation, Modify configuration, Alter behaviour)
Incident Status	Not solved
Total loss	£ 65,000

Table 20: Use Case 1 (CTI Dataset)

Associated Risk Evaluation

To compute the associated risk of sharing this CTI dataset, we apply our model as follows. The first step is to identify and analyse the severity for each property in the dataset. Table 21 defines the threats associated with disclosing the CTI dataset as derived from Table 20. We have assigned the sets of potential threats for each property and evaluated those for severity in cyber security contexts.

Property	**Property ID**	**Threat**	**Severity**
Victim (Name, Sector, Address, Role)	P1	T1, T2, T3, T4, T10	10
Malware (Type, Description)	P2	T3, T6	10
IoC (Name, Desciption, Value)	P3	T2, T3, T4, T5, T6	10
Vulnerability	P4	T2, T3, T4, T5, T6	10
Affected Assets (Type, Property)	P5	T2,T4,T7, T8, T9, T10	10
Status	P6	T2,T4,T6	10
Total Loss	P7	T6, T10, T11	50
Impact Assessment	P8	T6, T10, T11	10
Reporter	P9	T1, T2	10

Table 21: Severity value and Associated threats

Table 22 represents the same relationship between the threats and the properties of the CTI dataset by focusing on the threats.

Threat	**Threat ID**	**Matched Property**
Identity theft (Identity Fraud/ Account)	T1	P1, P9
Social engineering	T2	P1, P3, P4, P5, P6, P9
Unauthorised activities	T3	P1, P2, P3, P4
Targeted attacks (APTs etc.)	T4	P1,P3,P4, P5, P6
Misuse of information/ information systems	T5	P3, P4
Compromising confidential information (data breaches)	T6	P2, P3, P4, P6, P7, P8
Unauthorised physical access	T7	P5
Violation of laws or regulations / Breach of legislation	T8	P5
Failure to meet contractual requirements	T9	P5
Loss of reputation	T10	P1, P5, P7,P8
Judiciary decisions/court orders.	T11	P7 , P8

Table 22: Threats and matched property

Based on the CTI dataset disclosure and the associated threats we estimate the likelihood of a threat occurring based on the property value and the context which varies depending on the organisations' requirements.

Table 23 presents our estimates of the likelihood L_{ij} of the threats and the total risk score TAR when sharing with public sharing communities. In table 23, the risk level for each threat will be the likelihood of the threat and the severity of the exposure of the associated properties. From table 23, we can evaluate that the sub-risk value of (P1, P3, P4, P5, P6 and P9) properties when the threat is "Social Engineering" (T2). The total value is 28 using equation (6):
(1 * 10) + (0.1* 10) + (0.1 * 10) + (1 * 10) + (0.5 * 10) + (0.1 * 10) = 28
the sub-risk associated with sharing (P1, P3, P4, P5 and P6) properties when the threat is "Targeted attacks (APTs etc.)" (T4) is:
(1 * 10) + (0.5* 10) + (0.5 * 10) + (1 * 10) + (1 * 10) = 40.
The total risk will be adding all sub risk and the total value is 275

Table 24 presents the estimated likelihood of the threats and the total risk score value when sharing with trusted communities.
Finally, we evaluated the risk in three different scenarios: sharing the CTI dataset with public communities, sharing when involving/considering a high level of trust with the receiver and finally, sharing after removing the unrelated information.

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**SUB-RISK**
T1	0.1	0	0	0	0	0	0	0	0.1	2
T2	1	0	0.1	0.1	1	0.5	0	0	0.1	28
T3	0.5	1	0.5	0.5	0	0	0	0	0	25
T4	1	0	0.5	0.5	1	1	0	0	0	40
T5	0	0	0.1	0.1	0	0	0	0	0	2
T6	0	1	0.1	0.1	0	1	1	1	0	82
T7	0	0	0	0	0.5	0	0	0	0	5
T8	0	0	0	0	0.5	0	0	0	0	5
T9	0	0	0	0	0.5	0	0	0	0	5
T10	0.5	0	0	0	0.5	0	1	0.5	0	65
T11	0	0	0	0	0	0	0.1	0.1	0	6
TAR										**275**

Table 23: UC1 Likelihood and total risk value (public sharing communities)

	P1	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**SUB-RISK**
T1	0.1	0	0	0	0	0	0	0	0	1
T2	0.5	0	0.1	0.1	0.5	0.1	0	0	0.1	14
T3	0.1	0.5	0.5	0.5	0	0	0	0	0	16
T4	0.1	0	0.1	0.1	0.1	0.5	0	0	0	9
T5	0	0	0.1	0.1	0	0	0	0	0	2
T6	0	0.5	0.1	0.1	0	0.5	0.5	0.5	0	42
T7	0	0	0	0	0.1	0	0	0	0	1
T8	0	0	0	0	0.1	0	0	0	0	1
T9	0	0	0	0	0.1	0	0	0	0	1
T10	0.1	0	0	0	0.1	0	0.5	0.5	0	32
T11	0	0	0	0	0	0	0	0	0	0
TAR										**119**

Table 24: Likelihood and total risk value (trusted communities)

When sharing with public communities, the risk value is 275. On the other hand, sharing within trusted communities decreases the risk value to 119.
In this scenario, the purpose of sharing is to check the existence of the same malware thus we need to know the type and description of the malware, in addition to the indicators of compromise such as hash file value and windows registry key. Therefore, the properties needed for sharing are P2 and P3. Therefore, the associated risk value if we only share these essential properties will be reduced to 34 as shown in Table 25. Reducing the risk value is important for encouraging CTI sharing, and to achieve that, the organisation filters out the sensitive information that is not relevant to the purpose of this sharing.

	P2	**P3**	**SUB-RISK**
T1	0	0	0
T2	0	0.1	1
T3	1	0.5	15
T4	0	0.5	5
T5	0	0.1	1
T6	1	0.1	11
T7	0	0	0
T8	0	0	0
T9	0	0	0
T10	0.1	0	1
T11	0	0	0
TAR			**34**

Table 25: UC1 Likelihood and total risk value for sub-dataset

Our model allows for each risk assessment to be combined in different ways for different purposes. For instance, Figure 11 demonstrates a risk assessment visualisation for the same CTI dataset. For each field in the CTI dataset, we displayed the sum of the risks posed by that property in case of disclosure. This visualisation shows which properties of CTI datasets are the greatest risk when sharing and might be used in the context of raising organisational awareness of the CTI dataset fields.

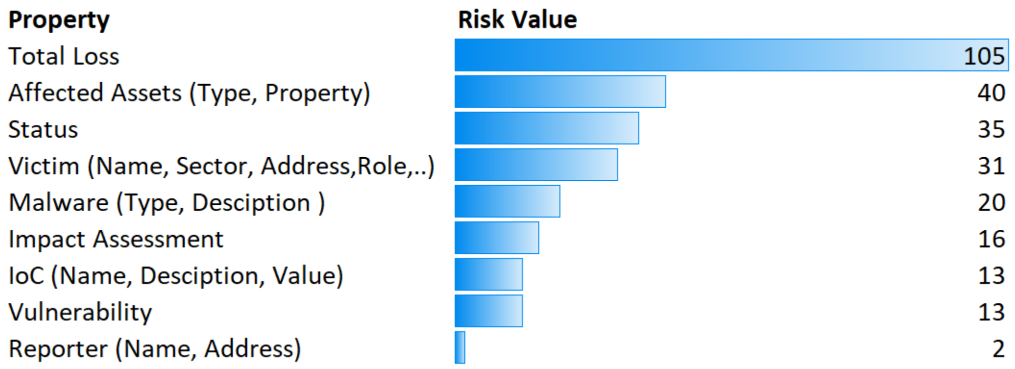


Figure 11: A risk assessment visualisation showing risk value per type of information

Evaluation - Data Collection and Analysis

This section presents the results of the data collection from a questionnaire, see Section 4.8, conducted within privacy and cybersecurity workshops with 15 experts in privacy and cybersecurity. The study provided anonymity to the participants. The questionnaire contains 3 parts. The first part focuses on identifying the threats associated with disclosing the CTI dataset. We proposed a list of threats and free text for extra suggestions. This part will validate our analysis of identifying the threats of disclosing sensitive and identifiable information in cyber incident information as proposed in Chapter 3. The second part focuses on the security controls that might be applied to preserve privacy of the dataset such as redaction/selection, anonymisation, aggregation, encryption, and so on. This part will give an insight of the required protection level and the technical methods that would help organisations to share CTI dataset and to ensure the confidentiality. Finally, the third part focuses on giving a risk value to the dataset in both cases, before and after applying the security controls. This part will help validate our ARM model.

Fifteen experts filled out the questionnaire, and a summary of the data collected is presented in Table 26 and discussed in more detail below.

Question	Part 1: sharing with public	Part 2: sharing with trusted entities
Q-1	15	12
Q-2	15	13
Q-3.1 (Redaction/Selection)	8	0
Q-3.2 (Anonymisation)	7	7
Q-3.3 (Aggregation)	6	7
Q-3.4 (Enc)	7	7
Q-3.5 (others)	3	3
Q4	14	14

Table 26: UC1 Summary: Responses Returned

Fifteen experts answered question Q1 for sharing the CTI dataset with public sharing community, and 12 experts answered the same question when sharing with trusted communities. Nine experts selected in detail the possible associated threats of disclosing this dataset. Table 27 presents the threats and how many experts have selected that threat as a possible threat in case of disclosing this CTI dataset. For example, six experts out of nine agreed that disclosing this dataset would be associated with “Compromising confidential information” and “Loss of reputation” threats. The remaining experts did not consider these as possible threats. To reduce the effect of experts’ subjectivity, we will measure the level of agreement between all opinions in addition to comparing them to our opinion. To find the level of agreements between our selection and the experts’ selection, we compute the Fleiss’ Kappa agreement score [213]. In this use case, we find that we have a “moderate” agreement level with six experts with k = 0.428 for data containing seven experts, including our own rating for 17 possible threats. Still, the rest of the experts agreed with some of the proposed threats. Therefore, the result indicates that the list we have proposed in Table 22 matches significantly with the experts’ selections in Table 27.

Threat	**Count**	**Threat**	**Count**
Social engineering (Phishing, Spear phishing)	4	Loss of reputation	6
Failure to meet contractual requirements	3	Unauthorised physical access	0
Violation of laws or regulations	2	Failed business process	2
Compromising confidential information (data breaches)	6	Man-in-the-middle / Session hijacking	0
Identity theft (Identity Fraud/ Account)	4	Terrorists attack	0
Abuse of authorisations	0	Targeted attacks (APTs etc.)	2
Misuse of information/ information systems	4	Unauthorized activities	4
Generation and use of rogue certificates	0	Manipulation of information	3

Table 27: UC1 Part1, Threat Summary

Table 28 presents the number of experts who decided which threats might be associated with disclosing the CTI dataset when sharing with trusted entities. The possible threats have decreased due to the increase of trust level among the sharing organisations. However, the result still shows that the list we have proposed in Table 22 matches the experts' selections in Table 28.

Threat	**Count**	**Threat**	**Count**
Social engineering (Phishing, Spear phishing)	5	Loss of reputation	6
Failure to meet contractual requirements	4	Unauthorised physical access	0
Violation of laws or regulations	1	Failed business process	2
Compromising confidential information (data breaches)	4	Man-in-the-middle / Session hijacking	0
Identity theft (Identity Fraud/ Account)	1	Terrorists attack	0
Abuse of authorisations	0	Targeted attacks (APTs etc.)	0
Misuse of information/ information systems	1	Unauthorized activities	0
Generation and use of rogue certificates	0	Manipulation of information	0

Table 28: UC1 Part2, Threat Summary

For question Q2, eight experts indicated that we cannot share this dataset. On the other hand, seven indicated that we could share after mitigation. This result indicates that sharing this dataset without applying any security controls will be a high risk to CyberA.

For questions Q3.1 and Q3.2, experts selected values that should be anonymised or removed from the dataset before sharing, such as "Reporter Name", "Reporter

Email”, “Reporter Address”, “Victim Name”, “Victim Sector”, “Victim Device”, “Victim Email”, “Victim Address“ and “Total Loss”.

Many experts agreed to remove any personal data, such as the victim information which will reduce possible threats such as “Violation of laws or regulations” and make the decision of sharing compliant with regulation such as the GDPR. In our model, we looked at the properties that will be useful for the purpose of sharing and the analysis as it is presented in Table 25. These fields are (Malware, observed-data, Indicator). Therefore, the experts’ selection is relevant to our model of risk value evaluation because the excluded properties will not be useful for the purpose of this sharing.

For question Q3.3, six experts gave an answer which included Address, Date and Affected Assets Type. This indicates that some information needs to be grouped and aggregated before sharing as part of reducing the risk of sharing individual information.

Also, sharing the full dataset would not be necessary to achieve the goal of this analysis, and it could reveal sensitive information which might be unimportant to other organisations and highly risky to share. Therefore, after evaluating the dataset we have extracted a sub-dataset which contains only the relevant information.

For question Q3.4, seven experts indicated that some attributes should be encrypted, such as indicator of compromise values, email addresses and victim information. This decision will work properly when CyberA needs to share the sub-dataset with other organisations where the level of trust is low and to avoid any inferring of sensitive information, such as network infrastructure from the network traces [214]. We can apply one of the several techniques to protect privacy in correlation, such as salted hashes [215] and homomorphic encryption [49]. By applying these techniques, an analyst can ask for a correlation and analysis without revealing extra information about what they are looking for.

For question Q3.5, three experts confirmed that specific fields such as IP addresses and email addresses should be generalised.

For question Q4.1, experts were asked to evaluate overall risk on a 1-5 scale, with 5 being the worst. Nine experts indicated that the risks are between 4 and 5 which constitutes a high level of risk. On the other hand, after applying the suggested controls, five experts suggested that the risk value would be between 1 and 2 which constitutes a low risk level. However, when sharing the CTI dataset with trusted entities, the overall value changed from a medium risk level to a low risk level. Eight experts stated that the risk value is between 2 and 3, and after applying the security controls, eight stated it was between 1 and 2.

As a result, the case study findings suggest that sharing this CTI dataset is possible after applying specific security controls, mainly by removing unrelated data and applying encryption. From the questionnaire results, we find out that our model reached an acceptable match with respect to the cybersecurity and privacy experts. All the threats we identified were also identified by the experts. Experts identified different controls to reduce the risk of sharing and they agreed that sharing this dataset without applying these controls is high risk. Although some experts had different decisions, this difference can be attributed to the different expertise levels and the experts' subjective view of how they define the granularity level of the risk. Also, threat and technical details such as network information can have different meaning between security experts. For example, five experts have not selected encryption as a security control which should have been applied before sharing, and others focused mainly on the anonymisation techniques as a security control. In our model, the dataset admin is free to select the security control of choice, for example, homomorphic encryption [49] [216] or Secure multiparty computation [217] [59].

4.5.2 Use Case 2: "CTI contains malware information & personal information – aggregation of data"

This scenario consists of cyber threat companies, CyberA and other companies, which share threat intelligence with one another. CyberA has been attacked by a specific threat actor and would like to know how many companies have been attacked by the same threat actor. Sharing the threat actor information is sensitive due to the possibility of identifying the techniques and procedures used in the attack, the victim information and the targeted sector such as oil business, health and diplomatic offices. The incentive of this sharing is to understand and analyse this threat actor. CyberA needs to determine how many companies have been targeted by the same threat actor.

In this case study, we have used the STIX report about the "Red October" Campaign [218]. Before sharing the STIX report, we need to evaluate the associated risk of sharing this information within the CTI sharing communities.

Table 29 shows the sample CTI dataset which contains the properties that might be shared.

Property	**Value**
TTP Malware Type	Command and Control, capture stored data, Scan network, Exploit vulnerability, Remote Access Trojan, Downloader, Export data, Spyware/Keylogger, Brute force
TTP Attack Patterns	CAPEC-98
Vulnerability	CVE-2009-3129, CVE-2008-4250, CVE-2010-3333, CVE-2012-0158, CVE-2011-3544
Title	Incident associated with Red October campaign. Phishing email with malware attachment leading to infection, C2, credential compromise, and lateral movement through the network.Goal to steal classified info and secrets
External ID	4F797501-69F4-4414-BE75-B50EDCF93D6B
Incident Date	2012-01-01T00:00:00
Reporter	Alex John, W-baker org, alex@w-baker.org, (LE1 9BH, Leicester, UK)
Victim	Japan Fair Trade Commission – intnldiv@jftc.go.jp
Victim Address	International Affairs Division (16th floor), Japan Fair Trade Commission, 6-B building, Chuo Godo Chosha, 1-1-1 Kasumigaseki, Chiyoda-ku Tokyo 100-8987
Affected Assets Type	Desktop, Mobile phone, Router or switch, Server, Person
Affected Assets Property	Confidentiality (Classified, Internal, Credentials, Secrets, System) Integrity (Software installation, Modify configuration, Alter behaviour)
Security Compromise	Yes
Discovery Method	Ext - suspicious traffic
Threat Actor Title	Lone Wolf Threat Actor Group
Threat Actor Description	Notes: Basing on registration data of command and control servers and numerous artefacts left in executables of the malware, we strongly believe that the attackers have Russian-speaking origins. Current attackers and executables developed by them have been unknown until recently, they have never related to any other targeted cyberattacks
Threat Actor	The Lone Wolf / Gookee Organisation
Threat Actor/ Country	Russia
Threat Actor/ Administrative Area	Moscow
Threat Actor Electronic /Address Identifier	lone-wolf@stealthemail.com / facebook.com/theLonewolf
Threat Actor Language	Russian
Threat Actor Motivation	Espionage
Threat Actor Observed TTPs	“example:ttp-fcfe52c2-3060-448b-b828-3e09341485b1” “example:ttp-2a884574-bf2b-4966-91ba-3e9ff6fea2e3” “example:ttp-22290611-0125-4c62-abcc-ddd4b8d3fb5d”

Table 29: UC2 Dataset

Associated Risk Evaluation

Analogous to use case 1, we have evaluated the associated risk of sharing the CTI dataset, we are applying our model as follows. Table 30 defines the threats associated with disclosing the CTI dataset and identifies the cybersecurity severity for each property as derived from Table 29.

Property	**Property ID**	**Threat ID**	**Severity**
TTPs	P1	T1, T2, T3, T9	50
Reporter	P2	T2, T4, T7, T9	10
Victim	P3	T2, T3, T4, T5, T6,T7, T9	50
Affected Asset	P4	T2, T3, T6, T7,T8 , T9	10
Threat Actors	P5	T1, T2, T3, T7, T9	50
Security Compromise	P6	T6	10
Discovery Method	P7	T6	10

Table 30: UC2 Associated threats and Severity value

Then we have Table 31, which represents Table 30 in a different way by focusing on the threats.

Threat	**Threat ID**	**Matched Property**
Compromising confidential information	T1	P1, P5
Social engineering	T2	P1, P2, P3, P4, P5
Targeted attacks (APTs etc.)	T3	P1, P3, P4, P5
Identity theft (Identity Fraud/ Account)	T4	P2, P3
Unauthorised activities	T5	P3
Loss of reputation	T6	P3, P4, P6, P7
Violation of laws or regulations / Breach of legislation	T7	P2, P3, P4, P5
Failure to meet contractual requirements	T8	P4
Misuse of information	T9	P1, P2, P3, P4, P5

Table 31: UC2 Threats and matched property

We estimate the likelihood of a threat occurring based on the property value and the context. For example, targeting high profile victims such as embassies will increase the probability of the "Misuse of information" threat in case of disclosing victim and attack vector information. The total associated risk (TAR) is the sum of sub associated risks of disclosing CTI information. Table 32 presents the likelihood L_{ij} of the threats and the total associated risk score TAR when sharing with public sharing communities. Table 33 presents the likelihood of the threats and the total risk score value when sharing with trusted communities.

Threat	**ID**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**SUB RISK**
Compromising confidential information	T1	1	0	0	0	1	0	0	100
Social engineering	T2	1	0.1	1	0.5	0.5	0	0	131
Targeted attacks (APTs etc.)	T3	0.5	0	0.5	0.5	0.5	0	0	80
Identity theft (Identity Fraud/ Account)	T4	0	0.1	0.1	0	0	0	0	6
Unauthorised activities	T5	0	0	0.1	0	0	0	0	5
Loss of reputation	T6	0	0	1	0.5	0	0.1	0.1	57
Violation of laws or regulations	T7	0	0.1	0.1	0.1	0.1	0	0	12
Failure to meet contractual requirements	T8	0	0	0	0.1	0	0	0	1
Misuse of information	T9	0.5	0.1	1	0.5	0.5	0	0	106
TAR									498

Table 32: UC2 Likelihood and total risk value (public sharing communities)

THREAT	**ID**	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**SUB RISK**
Compromising confidential information	T1	0.5	0	0	0	0.5	0	0	50
Social engineering	T2	0.5	0.5	0.5	0.1	0.1	0	0	61
Targeted attacks (APTs etc.)	T3	0.5	0	0.5	0.1	0.1	0	0	56
Identity theft (Identity Fraud/ Account)	T4	0	0.1	0	0	0	0	0	1
Unauthorised activities	T5	0	0	0	0	0	0	0	0
Loss of reputation	T6	0	0	0.5	0.1	0	0.1	0.1	28
Violation of laws or regulations	T7	0	0	0	0	0.1	0	0	5
Failure to meet contractual requirements	T8	0	0	0	0	0	0	0	0
Misuse of information	T9	0	0.1	0.1	0.1	0	0	0	7
TAR									**208**

Table 33: UC2 Likelihood and total risk value (trusted communities)

When sharing with public communities, the risk value is 498. On the other hand, sharing within trusted communities decreases the risk value by 58% making the value 208. To reduce the risk of sharing and preserve the privacy in the shared information, data minimisation should be applied to exclude sensitive information that is not relevant to the analysis from the original dataset. The sanitised dataset would fulfil the purpose and usefulness of sharing. In this use case, we keep two properties which are “TTPs” and “Threat_Actors”. The total risk score of the sub dataset after removing unrelated properties will be reduced to 280 as explained in Table 34.

Threat	ID	P1	P5	Sub Risk
Compromising confidential information	T1	1	1	100
Social engineering	T2	1	0.5	75
Targeted attacks (APTs etc.)	T3	0.5	0.5	50
Identity theft (Identity Fraud/ Account)	T4	0	0	0
Unauthorised activities	T5	0	0	0
Loss of reputation	T6	0	0	0
Violation of laws or regulations	T7	0	0.1	5
Failure to meet contractual requirements	T8	0	0	0
Misuse of information	T9	0.5	0.5	50
TAR				280

Table 34: UC2 Likelihood and total risk value for sub-dataset

Evaluation - Data Collection and Analysis

This section presents the result of the data collection using the same questionnaire as used for Section 4.5.1. Eleven experts filled the survey and a summary of the data collected is presented in Table 35 and discussed in more detail below.

Question	Part 1: sharing with public	Part 2: sharing with trusted entities
Q-1	11	10
Q-2	11	9
Q-3.1 (Redaction/Selection)	7	5
Q-3.2 (Anonymisation)	3	5
Q-3.3 (Aggregation)	3	1
Q-3.4 (Enc)	3	4
Q-3.5 (others)	0	0
Q4.1	9	9
Q4.2	6	6

Table 35: UC2 Analysis Summary: Responses Returned

The first question was answered by 11 experts for sharing the CTI dataset with public sharing communities and by 10 when sharing with trusted communities. Nine experts selected in detail the possible associated threats of disclosing this dataset. Table 36 presents the threats and how many experts selected that threat as a possible threat in case of disclosing this CTI dataset. For example, seven experts agreed that disclosing this dataset would be associated with “Compromising confidential information” and six experts agreed on “Social engineering” and “Loss of reputation” threats. The result indicates that the list we have proposed in Table 30 is very similar to the experts’ selections in Table 36. For example, we have not considered the “Man-in-the-middle” (MITM) threat. MITM relies on weakness of the communication between two components and is based on the report context and the dataset information. We found difficulty in envisaging this threat. Also, this threat was identified by only one expert. We calculate the Fleiss’ Kappa agreement score in this use case and we have a “moderate” agreement level with six experts with $k = 0.416$ for a data contains seven experts, including my rating for 18 possible threats.

Threat	**Count**	**Threat**	**Count**
Social engineering	6	Loss of reputation	6
Failure to meet contractual requirements	2	Unauthorised physical access	0
Violation of laws or regulations	4	Failed business process	2
Compromising confidential information	7	Man-in-the-middle	1
Identity theft (Identity Fraud/ Account)	3	Terrorist attack	1
Abuse of authorisations	1	Targeted attacks (APTs etc.)	5
Misuse of information	5	Unauthorised activities	3
Generation and use of rogue certificates	0	Manipulation of information	4

Table 36: UC2 Part1, Threat Summary

Table 37 presents the number of experts who decided which threats might be associated with disclosing the CTI dataset when sharing with trusted entities. As shown in Table 37 the set of possible threats has been reduced due to the increase of trust level among the sharing organisations. However, the result still shows that the list we have proposed in Table 30 is very similar to the experts' selections in Table 37.

Threat	**Count**	**Threat**	**Count**
Social engineering	1	Loss of reputation	5
Failure to meet contractual requirements	3	Unauthorised physical access	0
Violation of laws or regulations	2	Failed business process	1
Compromising confidential information	3	Man-in-the-middle	0
Identity theft (Identity Fraud/ Account)	1	Terrorists attack	1
Abuse of authorisations	0	Targeted attacks (APTs etc.)	1
Misuse of information	2	Unauthorised activities	1
Generation and use of rogue certificates	0	Manipulation of information	1

Table 37: UC2 Part2, Threat Summary

For question Q2, eleven experts indicated that we cannot share this dataset, or we can share after applying specific security controls. This result indicates that we need to apply security controls before sharing this dataset in order to reduce the risk of sharing.

For questions Q3.1 and Q3.2 experts select values that should be anonymised or removed from the dataset before sharing. Many of the experts propose that we need to remove all personal information and victim information such as the organisation's name. In this case the victim information is not related to the purpose of sharing which matches our model and evaluation.

For question Q3.3, three experts gave answers which included Address, Date and Affected Assets. This indicates that some information needs to be grouped and aggregated before sharing as part of reducing the risk of sharing individual information.

For question Q3.4, three experts indicate that some attributes should be encrypted, such as threat actor and TTPs information and we can use techniques that support operations on encrypted data such as homomorphic encryption and multiparty computation.

Finally, for question Q4.1, nine experts indicate that the risks are between 4 and 5 which constitutes a high level of risk. On the other hand, after applying the suggested controls, five experts suggest that the risk value would be between 1 and 2 which constitutes a low risk level. When sharing the CTI dataset with trusted entities, the overall value changed from a medium risk level to a low risk level. Eight experts state that the risk value is between 2 and 4, and after applying the security controls, six state that it is between 1 and 2.

Table 36 and Table 37 show that the number of selected individual threats in this use case is higher than the first use case. In addition, Table 32, Table 33 and Table 34 show that the total risk value of this use case is higher than the first use case risk value. This is rational due to the context of the second use case. The second use case is about an attack and threat actor targeting diplomatic institutions worldwide [219]. The threat actor developed their own malware for stealing sensitive information and used techniques such as valid accounts to get access to the victim network. From the questionnaire results we find that our model matches the experts' decisions. The risk value is high, so sharing this information publicly will put the organisation at a higher risk. Therefore, sharing this dataset with trusted communities or applying multiparty computation to get the analytics result will decrease the sharing risk.

4.5.3 Use Case 3: “Cyber threat intelligence contains malware information and personal information - sharing for detection”

This use case has been conducted within the AIR4ICS [220] project. The project develops a new agile incident response framework for industrial control systems.

The events in the project simulated a high pressure situation of a live cyber incident, cyber Red Team vs Blue Team scenarios. We have conducted the questionnaire during two events: the first event represented an infrastructure of UK deep seaport called CTI port. This includes the docking and berthing of ships, loading and unloading of cargo and the warehousing and distribution of goods via road, rail and sea. The port's systems consist of 3 main elements: an enterprise network, an operational technology component and ships systems.

In the first scenario, CTI Port has been attacked by specific malware. This malware was designed to steal encrypted files - and was even able to recover files that had been deleted. CTI Port wants to share this incident dataset with others in their sharing community and the board has agreed to share this report. The purpose of this sharing is to identify the threat actor behaviour and how they get in. Also, to check if the attacker is targeting specific business.

Table 38 shows the sample cyber threat intelligence dataset, which contains the properties that might be shared.

Property	**Value**
Incident Title	Incident associated with CTI port campaign. The main techniques are Brute force, credential compromise. The goal is to steal classified information and secrets.
Incident Category	Unauthorised Access (A group gains logical access without permission)
TTP Malware Type	Command and Control, capture stored data, Scan network, Exploit vulnerability , Remote Access Trojan, Downloader, Export data, Spyware/Keylogger, Brute force
Indicator of Compromise	File hash for malicious malware. This file hash indicates that a sample of malware alpha is present.
Indicator Value	Hashes:'SHA-256': 'ef537f25c895bfa7jfdhfjns73748hdfjkk5d789c2b76589fjfer8fjkdndkjn7yfb6c' Windows-registry-key: "HKEY_LOCAL_MACHINE\\SYSTEM\\ControlSet001\\Services\\MSADL3" IP: 147.228.151.30
Vulnerability	CVE-2009-3129, CVE-2008-4250, CVE-2010-3333, CVE-2012-0158, CVE-2011-3544
Incident Date	2019-09-25 10:18:00
Reporter	Alex John, W-baker org, alex@ctiport.csc, - (LE1 9BH, Leicester, UK)
Victim	CTI Port finance@ctiport.csc
Victim Address	CTI port - Main building (5th floor) - LE1 9BH, Leicester, UK
Affected Assets Type	CRM/Finance, Web server, Finance Lead WS, Accounting WS 192.168.125.112, 192.168.125.114, 192.168.125.129, 192.168.121.151
Affected Assets Property	Confidentiality (Classified, Internal, Credentials, Secrets, System) Integrity (Software installation, modify configuration, Alter behaviour)
Security Compromise	Yes
Discovery Method	Monitoring Service (This incident was reported by a managed security event monitoring service.)- suspicious traffic
Threat Actor Title	Lone Wolf Threat Actor Group
Threat Actor Description	Based on the registration data of CRM/Finance server and several pieces of evidence left in executables of the malware, we strongly believe that the attackers have Russian-speaking origins.
Threat Actor Org-Name	The Lone Wolf / Gookee Organisation
Threat Actor Country	Russia
Threat Actor Admin- Area	Moscow
Threat Actor E- Address Identifier	lone-wolf@stealthemail.com / facebook.com/theLonewolf
Threat Actor Language	Russian
Threat Actor Motivation	Espionage
Threat Actor Observed TTPs	"example: ttp-fcfe52c2-3060-448b-b828-3e09341485b1" "example:ttp-2a884574-bf2b-4966-91ba-3e9ff6fea2e3" IP address: 147.228.151.33 / 147.228.151.35
Course of Action	Block communication between the threat actor agents and the Finance/CRM Server. This server contains records about the shipments so there should be a high operational impact
Total Loss	75000£

Table 38: Use Case 3 (CTI Dataset)

Associated Risk Evaluation

Analogous to use case 1 and use case 2, we have evaluated the associated risk of sharing the CTI dataset, we are applying our model as follows. Table 39 defines the threats associated with disclosing the CTI dataset and identifies the cybersecurity severity for each property as derived from Table 38.

Property	**Property ID**	**Threat ID**	**Severity**
TTP Malware Type	P1	T2, T4, T6	50
Reporter information	P2	T1, T2, T3	10
Discovery Method	P3	T6, T7	10
Vulnerability	P4	T4, T6	50
Victim information	P5	T1, T2, T3, T4, T6, T7	50
Threat Actor information	P6	T2, T6, T4	50
Title	P7	T6	10
Affected Assets	P8	T1,T2, T4, T5, T8, T9	10
Course of Action	P9	T2,T6	10
Incident Date	P10	T2, T6, T8, T9	10
Security Compromise	P11	T6, T7	50
Total Loss	P12	T6, T7	50
Indicator of Compromise	P13	T2, T4, T6, T10	10
Incident Category	P14	T6	10

Table 39: UC3 Associated threats and Severity value

Then we have Table 40 which represents Table 39 in a different way by focusing on the threats.

Threat	**Threat ID**	**Matched Property**
Identity theft	T1	P2, P5,P8
Social engineering	T2	P1, P2, P5, P6, P8, P9, P10, P13
Unauthorised activities	T3	P2, P5
Targeted attacks (APTs etc.)	T4	P1, P4, P5, P6, P8, P13
Failed business process	T5	P8
Compromising confidential information	T6	P1, P3, P4, P5, P6, P7, P9, P10, P11, P12, P13, P14
Loss of reputation	T7	P3, P5, P11, P12
Violation of laws or regulations / Breach of legislation	T8	P8, P10
Failure to meet contractual requirements	T9	P8, P10
Misuse of information/ information systems	T10	P13

Table 40: UC3 Threats and matched property

We estimate the likelihood of a threat occurring based on the property value and the context. For example, targeting high profile victims such as critical infrastructure will increase the probability of "Misuse of the information" threat in case of disclosing victim and attack vector information. The total associated risk (TAR) is the sum of sub associated risks of disclosing CTI information. Table 41 presents the likelihood L_{ij} of the threats and the total associated risk score TAR when sharing with public sharing communities. Table 42 presents the likelihood of the threats and the total risk score value when sharing with trusted communities.

Threat ID	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**SUB RISK**
T1	0	0.1	0	0	0.5	0	0	0.1	0	0	0	0	0	0	27
T2	0.5	0.5	0	0	1	0.5	0	0.5	0.1	0.1	0	0	0.5	0	117
T3	0	0.1	0	0	0.5	0	0	0	0	0	0	0	0	0	26
T4	0.5	0	0	0.1	0.5	0.5	0	0.5	0	0	0	0	0.5	0	90
T5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	5
T6	0.1	0.1	0.5	0	0.5	0.5	0.1	0	0.5	0.1	0.5	1	0.5	0.1	149
T7	0	0	0.5	0	1	0	0	0	0	0	0.1	1	0	0	110
T8	0	0	0	0	0	0	0	0.5	0	0.1	0	0	0	0	6
T9	0	0	0	0	0	0	0	0.5	0	0.1	0	0	0	0	6
T10	0	0	0	0	0	0	0	0	0	0	0	0	0.5	0	5
														TAR	**541**

Table 41: UC3 Likelihood and total risk value (public sharing communities)

Threat ID	**P1**	**P2**	**P3**	**P4**	**P5**	**P6**	**P7**	**P8**	**P9**	**P10**	**P11**	**P12**	**P13**	**P14**	**SUB RISK**
T1	0	0.1	0	0	0.1	0	0	0.1	0	0	0	0	0	0	7
T2	0.1	0.1	0	0	0.5	0.1	0	0.1	0.1	0.1	0	0	0.1	0	40
T3	0	0.1	0	0	0.1	0	0	0	0	0	0	0	0	0	6
T4	0.1	0	0	0.1	0.1	0.1	0	0.1	0	0	0	0	0.1	0	22
T5	0	0	0	0	0	0	0	0.5	0	0	0	0	0	0	5
T6	0.1	0.1	0.5	0	0.5	0.1	0	0	0.1	0.1	0.5	1	0	0.1	119
T7	0	0	0.5	0	0.5	0	0	0	0	0	0.1	0.5	0	0	60
T8	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	1
T9	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	1
T10	0	0	0	0	0	0	0	0	0	0	0	0	0.1	0	1
														TAR	261

Table 42: UC3 Likelihood and total risk value (trusted communities)

When sharing with public communities, the risk value is 541. On the other hand, sharing within trusted communities decreases the risk value by 49% making the value 261. To reduce the risk of sharing and preserve the privacy in the shared information, data minimisation should be applied to exclude sensitive information that is not relevant to the analysis from the original dataset. The sanitised dataset would fulfil the purpose and usefulness of sharing. In this use case we keep the following properties which are “TTP Malware Type”, “Vulnerability”, “Threat

Actor information", "Affected Assets", "Indicator of Compromise" and "Incident Category". The total risk score of the sub dataset after removing unrelated properties will be reduced to 183 as explained in Table 43.

Threat ID	**P1**	**P4**	**P6**	**P7**	**P8**	**P13**	**P14**	**SUB RISK**
T1	0	0	0	0	0.1	0	0	1
T2	0.5	0	0.5	0	0.5	0.5	0	60
T3	0	0	0	0	0	0	0	0
T4	0.5	0.1	0.5	0	0.5	0.5	0	65
T5	0	0	0	0	0.5	0	0	5
T6	0.1	0	0.5	0.1	0	0.5	0.1	37
T7	0	0	0	0	0	0	0	0
T8	0	0	0	0	0.5	0	0	5
T9	0	0	0	0	0.5	0	0	5
T10	0	0	0	0	0	0.5	0	5
							TAR	183

Table 43: UC3 Likelihood and total risk value for sub-dataset

Evaluation - Data Collection and Analysis

This section presents the results of the data collection from a questionnaire, see Section 4.8, conducted within the AIR4ICS [220] project.

We have asked the red team to evaluate an incident report created based on the first event's scenario. All Red team members had several years of Red teaming, cyber incident response and cyber security experience. Eight experts filled out the questionnaire and a summary of the data collected is presented in Table 44 and discussed in more detail below.

Question	Part 1- Sharing with public (Number of responses)	Part2- sharing with trusted entities (Number of responses)
Q-1	8	8
Q-2	6	7
Q-3.1(Redaction/Selection)	8	8
Q-3.2 (Anonymisation)	6	8
Q-3.3 (Aggregation)	4	3
Q-3.4 (Enc)	0	1
Q-3.5(others)	0	0
Q4	8	8

Table 44: UC3 Responses Returned

All the experts answered Q1 for sharing the CTI dataset with both public sharing communities and trusted communities. All experts selected in detail the possible associated threats of disclosing this dataset. Table 45 presents the threats, and how many experts have selected that threat as a possible threat in case of disclosing this CTI dataset. For example, most of participants agreed that disclosing this dataset would be associated with “Social Engineering” and with “Loss of reputation” threat. To reduce the effect of experts' subjectivity, we will measure the level of agreement between all opinions besides our opinion. To find the level of agreement between our selection and the experts' selection, we compute the Fleiss' Kappa agreement score [213]. Kappa value evaluates the level of experts' agreement. The perfect agreement is when the Kappa value is 1. On the other hand, maximum disagreement value is 0. We find k= 0.417, which is considered “moderate” agreement for data contains nine experts, including my rating, to evaluate 16 threats.

The result indicates that the list we have proposed in Table 39 matches the experts’ selections in Table 45.

Threat	**Count**	**Threat**	**Count**
Social engineering (Phishing, Spear phishing)	7	Loss of reputation	7
Failure to meet contractual requirements	4	Unauthorised physical access	2
Violation of laws or regulations	4	Failed business process	1
Compromising confidential information	6	Man-in-the-middle / Session hijacking	0
Identity theft (Identity Fraud/ Account)	5	Terrorists attack	0
Abuse of authorisations	0	Targeted attacks (APTs etc.)	6
Misuse of information/ information systems	6	Unauthorized activities	3
Generation and use of rogue certificates	0	Manipulation of information	0

Table 45: UC3 Part1, Threat Summary

Table 46 presents the number of experts who decided which threats might be associated with disclosing the CTI dataset when sharing with trusted entities. The values have been changed significantly if cyber threat intelligence is shared with trusted entities. Most of the possible threats have been decreased with total 57% due to the increase of trust level among the sharing organisations. However, the result still shows that the list we have proposed in Table 39 matches the experts' selections in Table 46.

Threat	**Count**	**Threat**	**Count**
Social engineering (Phishing, Spear phishing)	2	Loss of reputation	7
Failure to meet contractual requirements	3	Unauthorised physical access	0
Violation of laws or regulations	4	Failed business process	1
Compromising confidential information	4	Man-in-the-middle	0
Identity theft (Identity Fraud/ Account)	0	Terrorists attack	0
Abuse of authorisations	0	Targeted attacks (APTs etc.)	3
Misuse of information/ information systems	1	Unauthorized activities	1
Generation and use of rogue certificates	0	Manipulation of information	0

Table 46: UC3 Part2, Threat Summary

For question Q2, three experts indicated that we cannot share this dataset. On the other hand, three experts indicated that we can share after applying the selected security controls. This result indicates that sharing this dataset without mitigation will be a high risk to CTI port. For question Q3.1, three experts selected what are the properties that should be shared to achieve the sharing goal. The properties include information such as TTP Malware Type, Vulnerability, Threat Actor information, Indicator of Compromise and Incident Category. Most of them agreed to remove Total loss, Victim information and Course of Action properties which matches our sanitised dataset. Most of the experts suggested anonymisation and aggregation as a security control that can be used for specific properties such as reporter information, affected assets and course of action, but none of them suggested encryption techniques. The reason could be that none of them were familiar with encryption techniques such as homomorphic encryption [49] [216] or Secure multiparty computation [217][59].

For question Q4.1, experts were asked to evaluate overall risk on a 1-5 scale, with 5 being the highest risk. Most of the experts' 75% indicated that the risks are between 5 and 4, which constitute a high level of risk. On the other hand, after applying the suggested controls, all experts suggested that the risk value would be between 2 or 1 which constitutes a low risk level. However, when sharing the CTI dataset with trusted entities, the overall value changed from a medium risk level to a low risk level. Four experts stated that the risk value is 3, and three stated that the risk value is 4, and after applying the security controls, all experts stated that the risk value is 2 or 1.

As a result, the case study findings suggest that sharing this CTI dataset is possible after applying specific security controls, mainly by removing unrelated data. From the questionnaire results we find out that our model reached a very high match with the cybersecurity experts. All the threats we identified were also identified by the experts. Experts identified different controls to reduce the risk of sharing and they agreed that sharing this dataset without applying these controls is high risk.

4.6 Threats to validity

In terms of the participants and sample size, 23 experts (3rd year PhD students, academics and industrial practitioners all working in cybersecurity) participated in this study where their feedback and evaluation were used to evaluate the model. The experts were introduced to the use cases they had in order to evaluate without a previous tutorial, so it is possible that the experts were not completely familiar with the cyber threat intelligence and cyber incident reports. We neither tracked the time of the evaluation nor created a controlled environment where experts are tracked more closely.

Concerning maturation, we have started with four use cases to be validated by each expert, but we noticed that the participants became tired and did not complete the full use cases. Therefore, we just used fifteen experts to validate the two use cases and eight experts to validate the third use case, which was conducted during the AIR4ICS project events. Also, adding up scalar ratings of individual risks does not in itself give meaningful numbers, and that as a consequence needs to be treated with caution. In this context, where a fixed set of risks has been identified to apply to the particular scenarios of incident information sharing, and all the risks are evaluated in each instance, there would be a little more value. Finally, concerning the generalisation, using academic and professional experts might help the generalisation of the results to be used in the industrial context. On the other hand, we might need more use cases to be able to generalise to real-world cyber threat intelligence platforms.

4.7 Conclusion

In this chapter, we present a new quantitative risk model for sharing CTI datasets. The main objective of this model is to develop a framework to support sharing decisions regarding which information to share, and with whom. We have extended our previous work, in Chapter 3 we performed a comprehensive analysis of incident reporting information through the STIX incident model to identify the threats of disclosing sensitive and identifying information. Here we have identified the potential threats associated with sharing a CTI dataset, computed the severity for each property, and we propose an estimation of the likelihood of the threats in case of property disclosure. Finally, we have calculated the total risk score of sharing a CTI dataset, and we addressed all risks associated with the data which will be shared. Based on the risk value, the organisations can select appropriate privacy preserving techniques to reduce the risk of sharing. In order to evaluate the model, we have asked experts' opinions for risk identification and evaluation for three different use cases.

4.8 Risk assessment questionnaire

Part1 - Sharing with open communities - Information can be shared and used without restriction, no rules or agreements to join these communities.

Q1- Which of the following can be possible risk(s) of disclosing this incident information? Please choose all that apply.

☐ Social engineering (Phishing, Spear phishing)
☐ Failure to meet contractual requirements
☐ Violation of laws or regulations
☐ Compromising confidential information (data breaches)
☐ Identity theft (Identity Fraud/ Account)
☐ Abuse of authorisations

□ Misuse of information/ information systems

□ Generation and use of rogue certificates

□ Loss of reputation

□ Unauthorised physical access

□ Failed business process

□ Man-in-the-middle / Session hijacking

□ Terrorists attack

□ Targeted attacks (APTs etc.)

□ Unauthorised activities

□ Manipulation of information

Other, please specify:

Q2 - Can this dataset be shared with this recipient in any form?

Q3 -If they can share this dataset, what controls and modifications:

Q3.1 - Removing/Selection [which data, please specify]

Q3.2-Anonymisation [of which fields]

Q3.3- Aggregation [of which fields]

Q3.4- Encryption [of which fields]

Q3.5- Other, please specify

Q4- In your opinion, what is the risk level of sharing this dataset in the following two cases:

Q4.1- Without proposed controls on a 1-5 (5=highest) scale?

Q4.2- With proposed controls on a 1-5 (5=highest) scale?

Part2 -Sharing with trusted communities – Dataset can be shared and used only with specific entities, usually they have specific rules or partnership agreement to join them.

Q1- Which of the following can be possible risk(s) of disclosing this incident information? Please choose all that apply.

□ Social engineering (Phishing, Spear phishing)
□ Failure to meet contractual requirements
□ Violation of laws or regulations
□ Compromising confidential information (data breaches)
□ Identity theft (Identity Fraud/ Account)
□ Abuse of authorisations
□ Misuse of information/ information systems
□ Generation and use of rogue certificates
□ Loss of reputation
□ Unauthorised physical access
□ Failed business process
□ Man-in-the-middle / Session hijacking
□ Terrorists attack
□ Targeted attacks (APTs etc.)
□ Unauthorised activities
□ Manipulation of information

Other, please specify:

Q2 - Can this dataset be shared with this recipient in any form?

Q3 -If they can share this dataset, what controls and modifications:

Q3.1 - Removing/Selection [which data, please specify]

Q3.2-Anonymisation [of which fields]

Q3.3- Aggregation [of which fields]

Q3.4- Encryption [of which fields]

Q3.5- Other, please specify

Q4- In your opinion, what is the risk level of sharing this dataset in the following two cases:

Q4.1- Without proposed controls on a 1-5 (5=highest) scale?

Q4.2- With proposed controls on a 1-5 (5=highest) scale?

Chapter 5

Sharing Cyber Threat Intelligence Under the General Data Protection Regulation

In this chapter[1], we consider how cyber intelligence sharing interacts with data protection legislation. Specifically, we present a model for sharing cyber threat intelligence under the GDPR. It is an approach for defining the required protection level on cyber threat intelligence datasets, if they contain personal data, as defined by the GDPR. Based on the GDPR rules, this approach would help to make the decision of sharing and processing personal information clear. Moreover, it helps to provide some practical and clear rules to build data sharing agreements between organisations, because during the evaluation phase, we establish the purpose of the sharing, the legal basis and security measures for compliance with the law. This chapter has two main contributions. First, to provide a decision process about sharing CTI datasets containing personal data in the context of the GDPR. Second, to convert existing legal grounds into rules that help organisations share

[1] This chapter is based on the conference paper "Sharing Cyber Threat Intelligence Under the General Data Protection Regulation" [221]

such data whilst being legally compliant with the GDPR. These rules establish an association between the CTI policy space and the defined protection levels. This chapter is divided into the following sections. Section 5.1 describes the steps of the methodology to build the approach. Section 5.2 gives several use cases of sharing CTI datasets to validate our approach. Section 6 summarises this chapter.

5.1 Methodology

This section presents the methodology we used to build an approach to evaluate the possibility of sharing personal data in the context of CTI datasets under the GDPR. The methodology consists of three main steps and is inspired by the DataTags project [222]. The first step is to define the possible levels of security requirements which agree with the principles considered by the GDPR when processing personal data in CTI datasets. The second step is to identify a *policy space*, i.e. a set of concepts, definitions, assertions and rules around the GDPR to describe the possible requirements for sharing CTI datasets. The last step is to build the decision graph, which defines the sequence of questions that should be traversed to establish and assess the legal requirements for CTI data sharing, represented with an outcome as so-called "tags". The DataTags project, developed by Latanya Sweeney's group at Harvard University, helps researchers and institutions to share their data with guarantees that releases of the data comply with the associated policy, including American health and educational legislation [223]. It consists of labelling a dataset with a specific tag based on a series of questions. Each question is created based on a set of assertions under the applicable policy.

5.1.1 Defining DataTags related to cybersecurity information sharing

The first step to achieving our goal is to define the tags that will be the possible decisions reached after a series of questions that interrogate CTI datasets for GDPR requirements. The legal requirements of the GDPR indicate in the first instance whether we can share or not. However, when the answer is positive, additional obligations for such sharing arise out of the principles and articles of the GDPR, in particular: the principle of data minimisation; the requirement that personal data must be processed securely; and that the data must not be retained when no longer relevant. Hence, the decision process also leads to conclusions on how sharing can take place by translating these constraints into technical requirements. All of this is represented in the "data tags" of the leaves of our decision graph. The organisations that are sharing CTI datasets should ensure that the receiving organisation understands the sensitivity of this information and receives clear instructions on what they are allowed to do with the information, e.g. potential on-sharing. We will follow the Traffic Light Protocol (TLP) [43] levels as a springboard, and expand them by adding security measures for each level in order to address the GDPR requirements of processing personal data when sharing CTI datasets. TLP was created to facilitate the sharing of information by tagging the information with a specific color. TLP has four colors, indicating different levels of acceptable distribution of data, namely [43]:

- WHITE - Unlimited.
- GREEN - Community Wide.
- AMBER - Limited Distribution.
- RED - Personal for Named Recipients Only.

This protocol records whether recipients may share this information with others. TLP protocol is used by CSIRT communities, Information Analysis and Sharing Centres (ISACS) and various industry sectors. This protocol is easy to use by tagging the dataset with a specific colour. Organisations have common understanding of these tags. That helps them to apply the TLP automatically without complex trainings and documentations. TLP simplicity makes it suitable for many real-world scenarios. However, it is not optimal for automated sharing, and it does not cover complicated situations. For example, a cyber incident report could be TLP: RED for all the receiving entities, except the sharer who can change the information, thus TLP: AMBER would be practical for the sharer [224]. We have extended this protocol by adding appropriate security measures that are required for the legality of CTI sharing. To increase the trustworthiness between the entities and to encourage entities to share CTI, we require the receiving organisation to apply these security measures whilst keeping in mind that, in general, organisations use different approaches and levels of security practices. However, enforcing the receiver to apply these security measures is a challenge in itself and is beyond the scope of this thesis, similarly to the enforcement of sticky policies as discussed in Section 2.3.5 . Table 48 shows the levels that we are going to use in order to label the shared datasets. Cells in columns “Type”, “Description”, and “Examples” are taken from the TLP description [225]. The values in columns “Security Measures” and “Transfer/Storage” are our proposals to meet the legislative requirements for securely sharing this data. We have proposed technical methods that would help organisations to achieve what the GDPR mandates as a technical requirement to ensure confidentiality and protect data subjects (Article 32). When proposing the security measures, we had to take into consideration with whom we are going to share CTI datasets and their trustworthiness, because recipients who cannot be relied upon to protect the shared information need to be eliminated from further sharing. We combine the notion of privacy preservation of the data with the trust level of the recipient organisation, and because of that, we recommend the use of the Attribute based Encryption (ABE) technique [62] [63]. For encryption, ABE can use any combination of a set of attributes as a public encryption key. Decryption privileges of the data in this type of encryption are not restricted to a particular identity but to entities with a set of attributes which may represent items such as business type and location. For example, an organisation chooses to grant access to an encrypted log of its internet traffic, but restricts this to a specific range of IP addresses. Traditional encryption techniques would automatically disclose the log file in case the secret decryption key is released. Table 47 lists example values of some attributes in the data. The first attribute is the location of the organisation. Due to the different legal systems associated with international transfer information exchange, we will consider three levels: National, EU and International. The second attribute is the sector of the organisation, because of the similarity of the working processes and procedures and likely similar threat models. The value might contain energy, health, education, finance and so on. Finally, the size of the organisation may be relevant because the number of employees has been empirically related to the number of threats [148]. To use ABE, before sharing the data with other organisations and in case it is not shared to the public, the Setup Key Authority generates a master secret key along with a public key. It publishes the public key so everyone has access to it. The key authority uses the master secret key to generate a specific secret key for the participating organisation in the sharing community. For example, there might be an organisation called “Alpha” which gets a specific secret key from the key generator authority. “Alpha” is an organisation operating at the national level in the telecom sector. Before sharing any dataset with “Alpha”, the user will encrypt the dataset that has its own specific access policy. Hence, this user encrypts the dataset such that anyone at the national level working with the telecom business will be able to decrypt it. The organisation sharing

Attribute	**Value**
Location	National, EU, Global
organisation sector/similarity of business	Central authority, similar business, connected groups, . . .
organisation size	Small, medium, big

Table 47: ABE attribute

CTI datasets generates ciphertext with this policy. As a result, the organisation "Alpha" will be able to decrypt the dataset.

At all levels, Green, Amber and Red, data will be encrypted using the ABE method. In addition, we need to consider the data minimisation principle as defined in GDPR Art.5(1)(c) "1. Personal data shall be: (c) adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (data minimisation)". Hence, sharing should be designed to provide only the required data to successfully achieve a specific goal. This implies that we should use the minimum amount of personal information to decrease any privacy risk on individuals whose personal data might be included. This corresponds with the approach in the case studies in Chapter 4, where we chose to share only the essential information. Doing so will reduce the risks of the following potential privacy attacks on the data:

Identity disclosure [226] [227]: this threat occurs when the attacker is able to connect a data subject with their record in a CTI dataset. For example, an attacker might identify a victim because the dataset contains direct identifying information such as an email address, IP address or credential information.

Membership disclosure [228]: this threat occurs when an attacker can derive that a specific data subject exists in the dataset. For example, the dataset contains information about specific malware victims. Any person established to be in the dataset reveals that this victim has been hacked by this malware.

Attribute disclosure [34]: this threat occurs when data subjects are linked with information about their sensitive attributes such as biometric data that is used

Type	Description	Examples	Security Measures	Transfer / Storage
WHITE	Information does not contain any personal data or sensitive information so it can be shared publicly.	Sharing public reports and notifications that give a better understanding of existing vulnerability.	Anonymization (Identity disclosure, Membership disclosure, Attribute disclosure).	Clear
GREEN	Information shared with community or a group of organisations but not shared publicly.	Sharing cybersecurity information within a close community. For example, sharing email with malware link targeting specific sector.	Anonymization (Identity disclosure) Attribute-Based Encryption (ABE)	Encrypted
AMBER	Share information with a specific organisation; sharing confined within the organisation to take effective action based on it.	Sharing cybersecurity information that contains indicators of compromise, course of action to a specific community or sector e.g. financial sector.	Anonymization (Identity disclosure) Attribute-Based Encryption (ABE)	Encrypted
RED	Information exclusively and directly given to single identified party. Sharing outside is not legitimate.	Sharing that you have been attacked or notifying central authority about an incident.	Attribute-Based Encryption (ABE). Data minimisation to share only relevant data.	Encrypted

Table 48: Proposed DataTags relating to four proposed classes of access

to uniquely identify an individual. Some personal information is more sensitive and defined as “special category” under the GDPR. The GDPR (Art. 9) defines special categories that need extra protection and prohibits processing this type of data unless certain conditions apply.

There are methods to remove personal information from an individual’s record in a way that decreases the possibility of all these attacks, as described in more details in Section 2.3. Some of these methods that we can use are k-anonymity [19] which uses suppression and generalization as the main techniques, l-diversity [34] which is an extension of k-anonymity to protect the shared data against background knowledge and Homogeneity Attacks, and t-closeness [35] which is another extension of l-diversity that decreases the granularity and makes the distribution of the sensitive attribute close to the distribution of the entire attribute.

5.1.2 Policy space

We build the policy space of our model as a set of assertions using the context of the CTI dataset. The evaluation of cases will be based on the defined assertions.

The assertions will contain the legal grounds under which personal data can be processed, in this case for the purpose of ensuring network and information security. For instance, assertions for sharing CTI information with other parties are based on both the purpose of sharing which is “GDPR Recital 49 - ensuring network and information security” such as the prevention of any access to the critical system after credentials leaks, and the related legal basis which is “GDPR Art 6.1 (c) - processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party”. These steps offer a clear, practical framework, justifying the sharing of Cyber Threat Intelligence. The tagged data which meets the rules based on applicable assertions will be derived from the decision graph. In order to build the CTI policy space, we use a JSON file maintained by Computer Incident Response Center Luxembourg CIRCL [225] for the related context of use of data by CSIRTs. The goal of the file is to track processing personal information activities and support automation. Many assertions refer to GDPR Art. 30 which prescribes all the recordable details of processing activities. The main categories of the assertions contain:

- Purpose: “The purpose of the processing. Ref GDPR Art. 30 (1) (b)”, for example, “GDPR Recital 49 - the processing of personal data to the extent strictly necessary and proportionate for the purposes of ensuring network and information security”

- Legal ground: “Lawfulness/grounds for the processing activity. Ref GDPR Art. 6 & 5 (a).”, for example, “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party”.

- Data subjects: “Categories of the data subjects. Reference GDPR Art. 30 (1) (c).”.

- Personal data: “Personal data processed. Reference GDPR Art. 30 (1) (c).”, for example, information extracted from computer and networking systems.

- Recipients: “Categories of recipients. Reference GDPR Art. 30 (1)(d).” for example, suppliers and government institutions.
- International transfer: “Whether any personal data in this processing activity is transferred to a third country or an international organisation. Reference GDPR Art. 30 (1) (e)”. for example, “Transfer of data is required by a legal entity in a third country and is based on an international agreement in force between the requesting third country and the Union or a Member State (GDPR Art. 48)”.
- Retention period: “Retention schedule/storage limitation. Reference GDPR Art. 30 (1) (f) and Art. 5 (e)”.
- Security measures: “Security measures & Integrity & Confidentiality. Security measures can be technical and/or organisational. Reference GDPR Art. 30 (1) (e), 32 (1) and Art. 5 (f).”, for example, pseudonymisation.

Based on the previous assertion list, we need to extract the relevant assertion categories specifically related to CTI sharing. We will consider only those assertions that are directly related to CTI sharing. In the GDPR the purpose of processing personal data should be precise and for that the GDPR offers clear recognition of “ensuring network and information security” GDPR Recital 49 as the purpose of processing personal data for actors such as public authorities and CSIRTs. The legal grounds for processing personal data are provided in GDPR Art. 6 & 5 (a). CIRCL has published a discussion [229] of the legal grounds of information leak analysis and the GDPR context of collection, analysis and sharing information leaks. The legal grounds relevant in our context are “processing is necessary for the compliance with a legal obligation to which the controller is subject” where it applies to CSIRTs and data protection authorities and “processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party” otherwise. In the “legitimate interest” sharing CTI information will enable organisations to better detect and prevent attacks by, for example, identifying the IP address of a malware communications and control hub. We do not consider "consent" GDPR Art. 6 (1) (a) a credible legal basis for processing personal data in the context of sharing cyber threat Intelligence. This is because it is very hard to get consent of data subjects especially when dealing with huge amounts of data [229] (e.g. 1bn Yahoo accounts were compromised from a 2013 hack [230]) or when personal data such as IP addresses concerns the perpetrator of a cyber-attack. Also, vital interest Art.6(1)(d) is not feasible to be used to justify sharing and processing CTI, as there is no personal data in CTI datasets which would relate to a threat to life. However, the public interest Art.6(1)(e) would be the justification to process personal data in the case of acting under specific authorization from an official authority to check that the cyber incident could affect the public interest. The description of the personal data that pertains directly to the GDPR is described in Art.30(1)(c). The conditions under which personal data can be transferred to third countries or an international organisation are described in GDPR Art.30(1)(e). As a result, the CTI policy space is described in Figure 12.

5.1.3 Decision graph

In this step, we propose an assessment based on the previous assertions. This assessment contains a set of questions, and the answer to each question will lead to different questions or a final decision and as a result, we will assign a specific tag to the CTI dataset or even in some cases, the decision would be to not share. This assessment is not definitive, but it gives a chance to reflect on our understanding of sharing CTI datasets under the GDPR. Figure 13 shows the decision graph for sharing CTI datasets under the GDPR. Some of the decisions in the graph still require human judgement, so we make no claims of the process being fully automatable. This judgement could be assisted by the Data Protection Officer

(DPO) whose main duties are ensuring compliance with the GDPR and providing support regarding data protection (Article 37) (Recital 97). The GDPR requires the appointment of a DPO in a public authority or organisations performing specific risky types of processing actions (Article 37) (Recital 97). The process first establishes whether the proposed data sharing falls within the scope of the GDPR. Then it establishes the legal basis for any special category data included. This is likely to be rare in CTI datasets, but we could imagine biometric data following an attack that included a physical breach. Next, it establishes the legal basis for the overall processing. Then, it checks and selects appropriate retention and security protections. We assume the “trust level” node’s value has been determined based on previous knowledge of the trustworthiness of the entity that we are looking to share with. The outcome matches one of the TLP tags as described in the previous section. Of course, the CTI datasets are also likely to contain “sensitive” information about the infected asset and the exploitable vulnerability that should be protected as discussed in detail in Chapter 3. The outcome reflects concerns for the data protection angle only; included information that is sensitive in a different dimension might independently require strengthening of the security measures.

5.2 Use cases

Sharing information regarding current or ongoing attacks including information on threat actors, attack vectors, victims and impact of the attack is an essential scenario of sharing cyber threat intelligence. In order to see how to apply the tags on CTI datasets three different use cases were developed. In the first use case, the organisation that is the victim informs a central authority about the attack. In the second use case, an organisation informs another organisation about a recent attack that affects the availability, confidentiality or integrity of services. In the

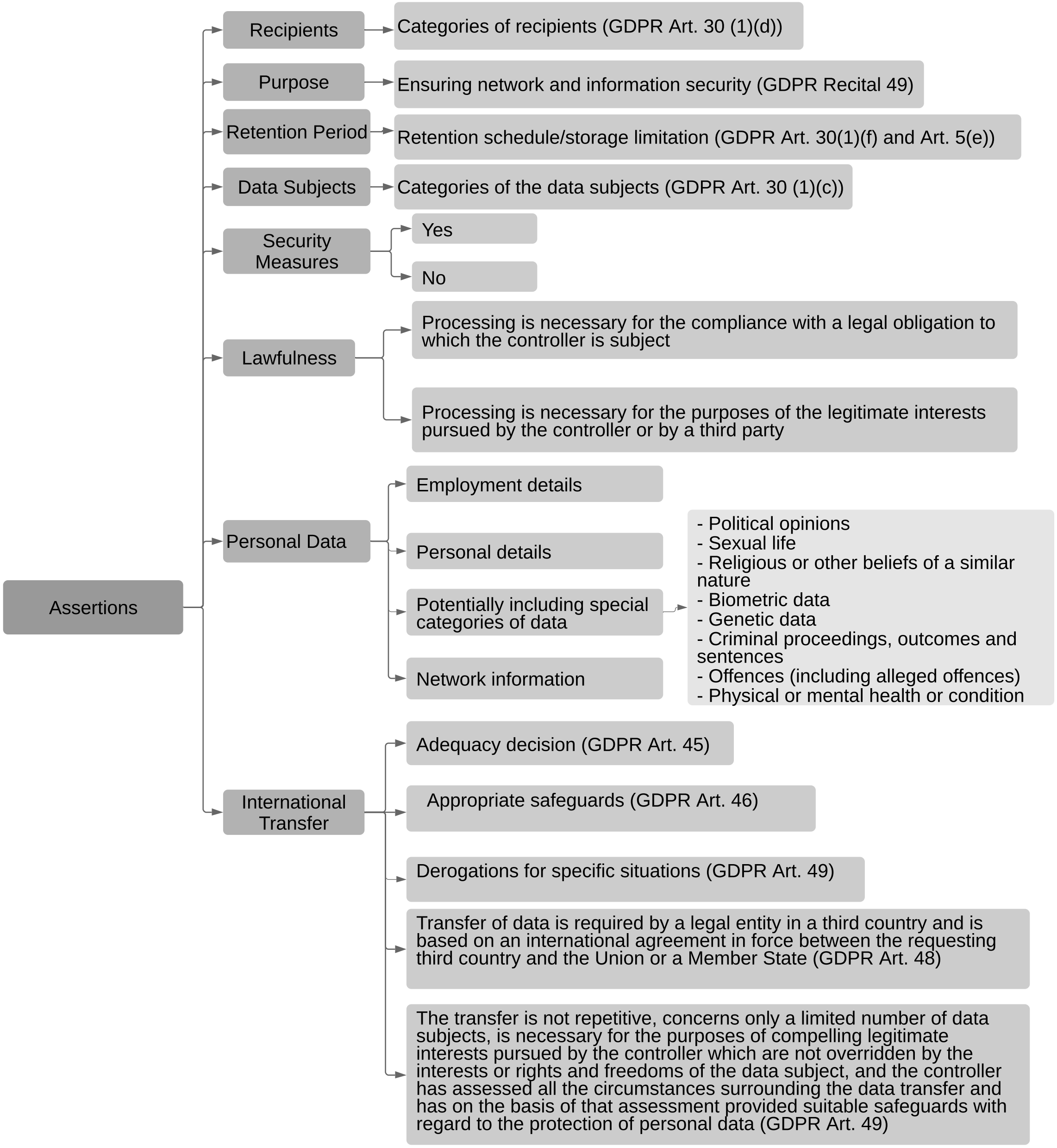

Assertions
Recipients
Categories of recipients (GDPR Art. 30 (1)(d))
Purpose
Ensuring network and information security (GDPR Recital 49)
Retention Period
Retention schedule/storage limitation (GDPR Art. 30(1)(f) and Art. 5(e))
Data Subjects
Categories of the data subjects (GDPR Art. 30 (1)(c))
Security Measures
Yes
No
Lawfulness
Processing is necessary for the compliance with a legal obligation to which the controller is subject
Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party
Personal Data
Employment details
Personal details
Potentially including special categories of data
Network information
- Political opinions
- Sexual life
- Religious or other beliefs of a similar nature
- Biometric data
- Genetic data
- Criminal proceedings, outcomes and sentences
- Offences (including alleged offences)
- Physical or mental health or condition
International Transfer
Adequacy decision (GDPR Art. 45)
Appropriate safeguards (GDPR Art. 46)
Derogations for specific situations (GDPR Art. 49)
Transfer of data is required by a legal entity in a third country and is based on an international agreement in force between the requesting third country and the Union or a Member State (GDPR Art. 48)
The transfer is not repetitive, concerns only a limited number of data subjects, is necessary for the purposes of compelling legitimate interests pursued by the controller which are not overridden by the interests or rights and freedoms of the data subject, and the controller has assessed all the circumstances surrounding the data transfer and has on the basis of that assessment provided suitable safeguards with regard to the protection of personal data (GDPR Art. 49)

Figure 12: CTI Policy Space

third use case, an organisation reports a security breach to the central authority.

5.2.1 Use Case 1: Informing central authority

This case study consists of two organisations, A and C (Central Authority) where an organisation A wants to report an incident to organisation C about a remote access tool (RAT) used by different threat actors. Before sharing the information, the reporter wants to be sure that sharing it is legitimate under the GDPR.

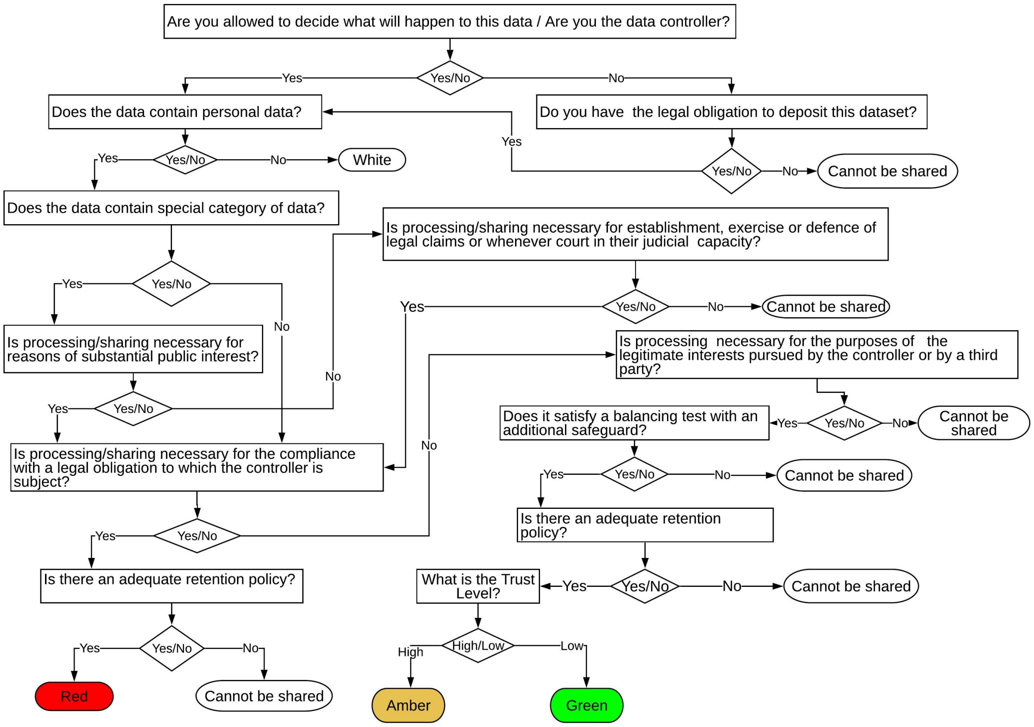

Are you allowed to decide what will happen to this data / Are you the data controller?
Yes/No
Yes
No
Does the data contain personal data?
Do you have the legal obligation to deposit this dataset?
Yes
Yes/No
No
White
Yes/No
No
Cannot be shared
Does the data contain special category of data?
Is processing/sharing necessary for establishment, exercise or defence of legal claims or whenever court in their judicial capacity?
Yes
Yes/No
Yes
Yes/No
No
Cannot be shared
Is processing/sharing necessary for reasons of substantial public interest?
No
Is processing necessary for the purposes of the legitimate interests pursued by the controller or by a third party?
No
Yes
Yes/No
Does it satisfy a balancing test with an additional safeguard?
Yes
Yes/No
No
Cannot be shared
No
Is processing/sharing necessary for the compliance with a legal obligation to which the controller is subject?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
Is there an adequate retention policy?
Is there an adequate retention policy?
What is the Trust Level?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
No
High
High/Low
Low
Red
Cannot be shared
Amber
Green

Figure 13: Decision graph

Discussion: The incident report contains personal information such as contact information of the reporter and credential information. Therefore, sharing and processing of such personal data would need to be legitimate under the GDPR. In order to decide how to share this information, the reporter needs to run an evaluation. The organisation A is the owner of this dataset and has the right to process this information, hence in this scenario the organisation A is considered the controller. Although the incident information contains personal data, it does not contain any special category data, such as, biometrics or political opinion, religious or philosophical beliefs, etc. In order to share this information with a Computer Security Incident Reporting Team CSIRT or the central authority, the reporter can rely on GDPR Art. 6 (1) (c) where the legal ground states “processing is necessary for the compliance with a legal obligation to which the controller is subject”. Organisation A has a retention policy in place. The security measures that should be applied to reduce the risk of harm to data subjects before sharing this dataset are: encrypted storage associated with a secure protocol to transmit this information. Moreover, the data will be encrypted by using ABE techniques with the properties (National, CA, Big) so as a result the final tag for this data will be RED. Figure 14 shows a sample questionnaire covering this case study.

5.2.2 Use Case 2: Sharing information about port scanning for incident prevention

This scenario consists of an organisation O1 in the energy sector which detects port scanning from a specific IP address for port range 0–1023 which is considered a potential threat. For incident prevention purposes, they may want to share information containing the source IP address, port range, the time of the incident, signs of the incident, and the course of action such as improve monitoring on these ports. The personal information in this scenario consists of the reporter information along with that of the individual who has made the observations, plus IP addresses which may be personal information of the attacker.

Discussion: organisation O1 is the controller of this data and needs to share

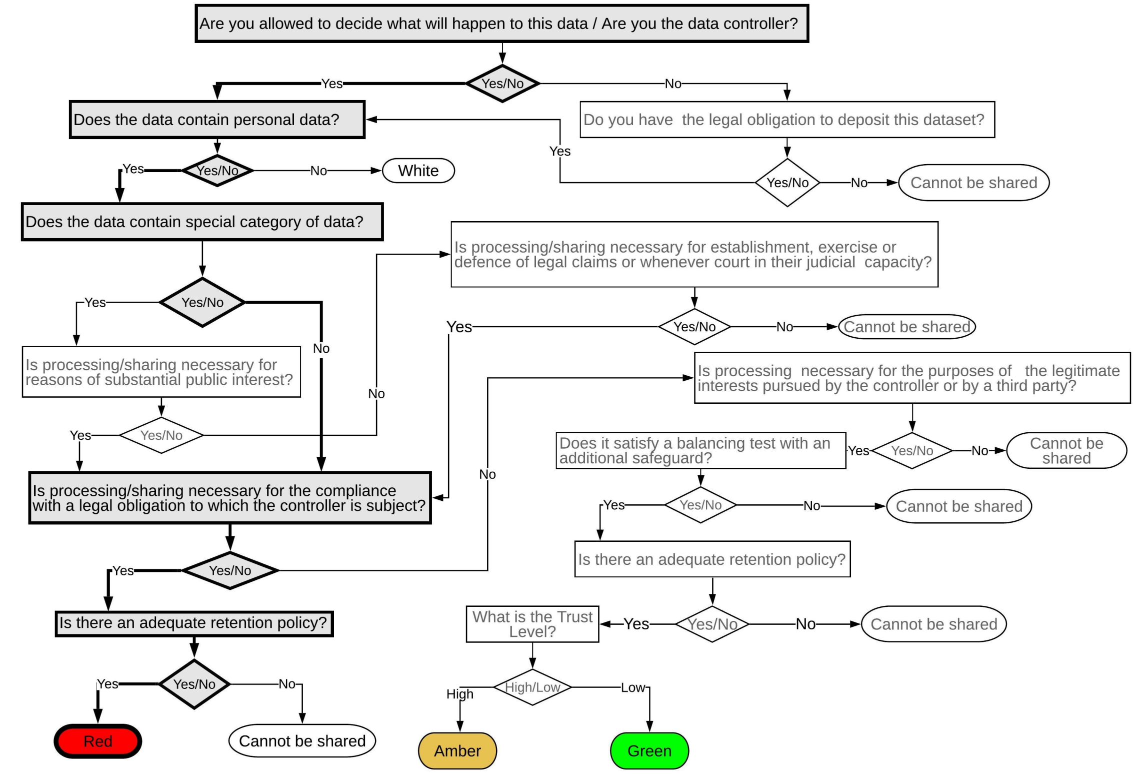

Are you allowed to decide what will happen to this data / Are you the data controller?
Yes/No
Yes
No
Does the data contain personal data?
Do you have the legal obligation to deposit this dataset?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
No
White
Does the data contain special category of data?
Is processing/sharing necessary for establishment, exercise or defence of legal claims or whenever court in their judicial capacity?
Yes
Yes/No
No
Is processing/sharing necessary for reasons of substantial public interest?
Yes
Yes/No
No
Cannot be shared
Is processing necessary for the purposes of the legitimate interests pursued by the controller or by a third party?
Yes
Yes/No
No
Does it satisfy a balancing test with an additional safeguard?
Yes
Yes/No
No
Cannot be shared
Is processing/sharing necessary for the compliance with a legal obligation to which the controller is subject?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
Is there an adequate retention policy?
Is there an adequate retention policy?
What is the Trust Level?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
No
High
High/Low
Low
Red
Cannot be shared
Amber
Green

Figure 14: Use case 1 decision graph (following the bold lines)

this information with trusted company O2. Because the dataset contains personal information, sharing needs to be legitimate under the GDPR. The dataset does not contain any special category data so we can continue and check the purpose of this sharing, which is GDPR Recital 49 – "ensuring network and information security". The reporter can rely on GDPR Art. 6(1)(f). The legal ground for sharing this information is "processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party". Presumably there is a retention policy in place. The security measures that will be associated before sharing this dataset are: encrypted storage associated with a secure protocol to transmit this information, anonymisation of reporter information against any identity disclosure and the data will be encrypted by using ABE techniques associated with the properties (EU, Energy sector, Medium).By applying these controls, the shared

CTI dataset satisfies the data minimisation rules. The trust level based on an assumed external calculation is high so as a result the final tag for this data will be AMBER. Figure 15 shows a sample questionnaire covering this case study.

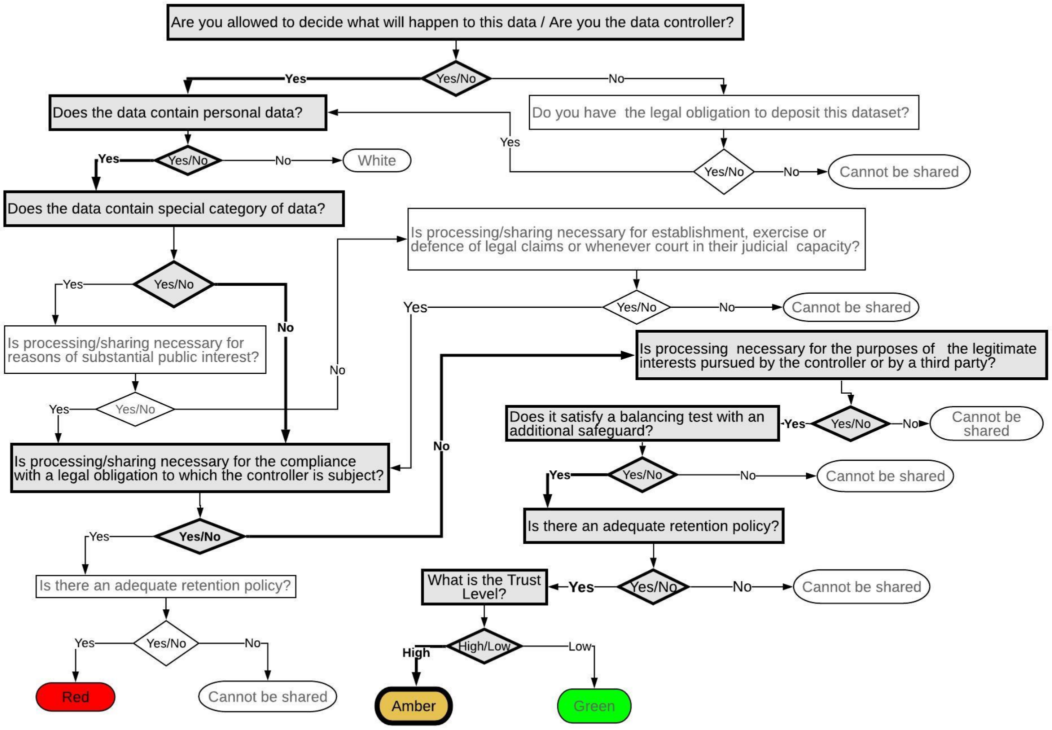

Are you allowed to decide what will happen to this data / Are you the data controller?
Yes
Yes/No
No
Does the data contain personal data?
Do you have the legal obligation to deposit this dataset?
Yes
Yes
Yes/No
No
White
Yes/No
No
Cannot be shared
Does the data contain special category of data?
Is processing/sharing necessary for establishment, exercise or defence of legal claims or whenever court in their judicial capacity?
Yes
Yes/No
No
Yes
Yes/No
No
Cannot be shared
Is processing/sharing necessary for reasons of substantial public interest?
Is processing necessary for the purposes of the legitimate interests pursued by the controller or by a third party?
No
Does it satisfy a balancing test with an additional safeguard?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
No
Yes
Yes/No
No
Cannot be shared
Is processing/sharing necessary for the compliance with a legal obligation to which the controller is subject?
Is there an adequate retention policy?
Yes
Yes/No
What is the Trust Level?
Yes
Yes/No
No
Cannot be shared
Is there an adequate retention policy?
High
High/Low
Low
Yes
Yes/No
No
Red
Cannot be shared
Amber
Green

Figure 15: Use case 2 decision graph (following the bold lines)

5.2.3 Use Case 3: Sharing CTI incident report for legal obligation

This scenario will cover reporting a security breach on organisation Alpha to the central authority. The incident covers the following “Sensitive information belonging to jobseekers has been put at risk on the government's new Universal Jobmatch website, it has been reported. The security flaw was uncovered during a Channel 4 News investigation. Hackers were said to have been able to register as an employer on the site which is accessed through the Gov.uk portal – another

Property	Value
Title	“Sensitive information belonging to jobseekers has been put at risk on the government's new Universal Jobmatch website, it has been reported. The security flaw was uncovered during a Channel 4 News investigation. Hackers were said to have been able to register as an employer other site which is accessed through the Gov.uk portal – another website that has just been launched by the government to deliver morepublic services online. The hackers were reported to have obtained information including passwords and passport and driving licence scans after posting a fake advert for a cleaner on Universal Jobmatch” [231]
Initial_Compromise	2012-01-01T00:00:00
Reporter Name	Alex John
Reporter Affliliation	LLC
Reporter Email	alex@llc.co.uk
Reporter Addresses	GB-London
Victim Name	Universal Jobmatch
Victim Addresses	GB-London
Affected_Assets	Web application
Property_Affected	Confidentiality (Personal Information)
Impact_Qualification	Painful
Leveraged_TTP	Used Malware
Security_Compromise	Yes
Discovery_Method	Agent Disclosure (This incident was disclosed by the threat agent (private blackmail).
Threat_Actors	DarkHydruz
Threat actor description	DarkHydrus [232] is a threat group that has targeted multiple victims including government authorities and educational institutions in the Middle East since at least 2016. The group uses open-source tools and custom payloads for achieving successful attacks.
Threat actor Motivation	Financial or Economic

Table 49: UC3 - Sample of the Cyber Incident Report

website that has just been launched by the government to deliver more public services online. The hackers were reported to have obtained information including passwords and passport and driving licence scans after posting a fake advert for a cleaner on Universal Job-match.” [231] We have updated the report and completed the values of the STIX incident report. The new report contains personal information such as reporter name, email address and victim information. In addition, it contains several sensitive properties such as the impact assessment value is “Painful” which means that this incident has a real critical effect on the business process. The victim information which is an official website will lead to loss of reputation. The initial compromise that tells us when the attack has been discovered and more forensic investigation will provide information on how long the attack has existed in the attacked system. There was not any detailed information about the threat actors other than the location and the motivation, but this information may reveal extra information about the techniques that were used and the targeting victims. Table 49 shows the sample of the cyber incident report which contains the properties that the reporter wants to share.

Discussion: The incident report contains personal information. Therefore, sharing and processing of such personal data would need to be legitimate under the GDPR. In order to decide how to share this information, the reporter needs to run an evaluation. The organisation Alpha is the owner of this dataset and has the right to process this information, hence in this scenario the organisation Alpha is considered the controller. Although the incident information contains personal data, it does not contain any special category data, such as, biometrics or political opinion, religious or philosophical beliefs, etc. In order to share this information with the central authority, the reporter can rely on GDPR Art. 6 (1) (c) where the legal ground states "processing is necessary for the compliance with a legal obligation to which the controller is subject". organisation Alpha has a retention policy in place. The security measures that should be applied to reduce the risk of harm to data subjects before sharing this dataset are: encrypted storage associated with a secure protocol to transmit this information. Moreover, the data will be encrypted by using ABE techniques with the properties (National, CA, Big) so as a result the final tag for this data will be RED. Figure 16 shows a sample questionnaire covering this case study.

As a result, we present three use cases for sharing CTI datasets between different entities. The datasets have been evaluated based on the decision graph built in Section 5.1.3. The decision is positive in all use cases, but it is associated with different protection levels based on the flow of the assertions. Hence, our approach can give any organisation which intends to share CTI datasets the ability to determine that they are legally compliant with the GDPR.

5.3 Conclusion

In this chapter, we have presented an approach that can help different entities to make a decision compliant with the GDPR when sharing CTI datasets. We

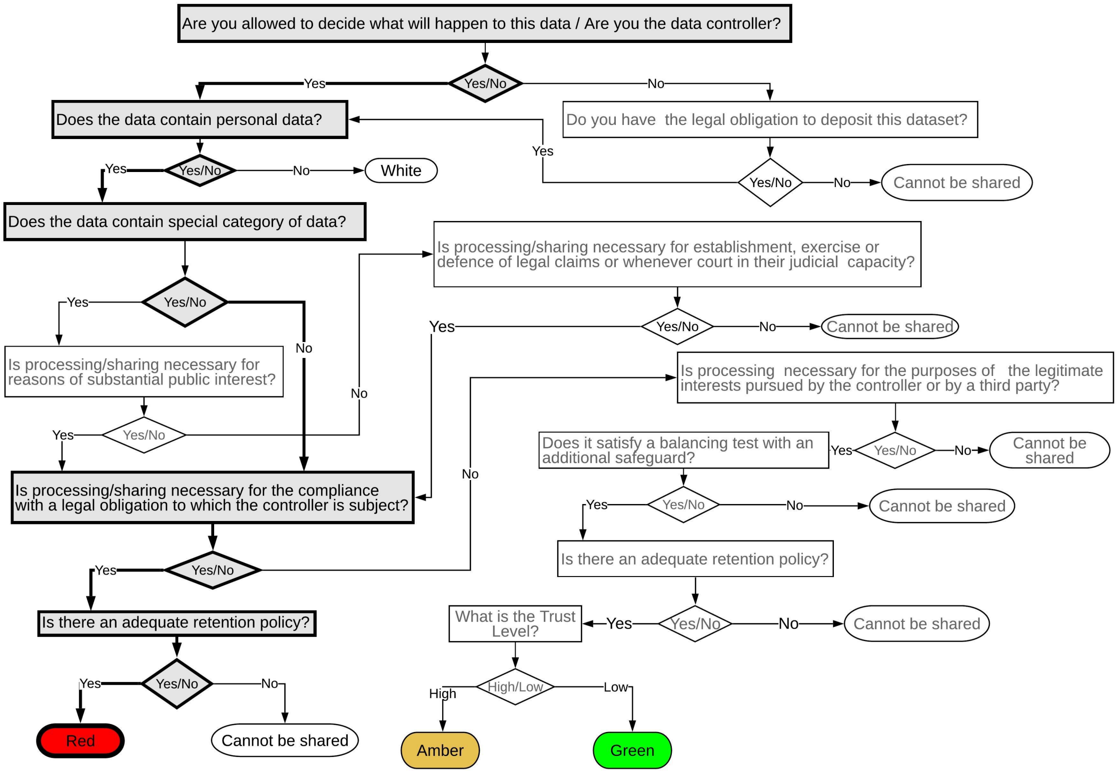

Are you allowed to decide what will happen to this data / Are you the data controller?
Yes/No
Yes
No
Does the data contain personal data?
Do you have the legal obligation to deposit this dataset?
Yes
Yes/No
Yes
No
White
Yes/No
No
Cannot be shared
Does the data contain special category of data?
Is processing/sharing necessary for establishment, exercise or defence of legal claims or whenever court in their judicial capacity?
Yes/No
Yes
No
Yes
Yes/No
No
Cannot be shared
Is processing/sharing necessary for reasons of substantial public interest?
Is processing necessary for the purposes of the legitimate interests pursued by the controller or by a third party?
No
Yes
Yes/No
Does it satisfy a balancing test with an additional safeguard?
Yes
Yes/No
No
Cannot be shared
No
Is processing/sharing necessary for the compliance with a legal obligation to which the controller is subject?
Yes
Yes/No
No
Cannot be shared
Is there an adequate retention policy?
Yes
Yes/No
Is there an adequate retention policy?
What is the Trust Level?
Yes
Yes/No
No
Cannot be shared
Yes
Yes/No
No
High
High/Low
Low
Red
Cannot be shared
Amber
Green

Figure 16: Use case 3 decision graph

have suggested adequate privacy preserving methods that should be applied when sharing CTI datasets. Then we have defined the policy space that related to the CTI in the context of the GDPR and finally built the decision graph that checks the legal requirements and provides a decision on how to share this information. There are limitations to our approach. In complex use cases, the decisions in the assessment graph may still be very demanding, such as whether the Recital 49 objective justifies any privacy impacts on the data subject. Furthermore, including additional regulations or local policies besides the way they will interact with the GDPR requirements would make the decision graph more complex. Additional legal and technical requirements might make the data tag collection harder to structure and manage, as well as complicating the decision process. In chapter 3, we have identified the associated threats of disclosing CTI. Here we have specifically addressed the legal risks associated with sharing CTI datasets. Our overall work aims to mitigate all threats associated with sharing CTI datasets and improve the sharing process.

Chapter 6

Conclusion

Sharing cyber threat intelligence may help organisations to better protect themselves against future cyber attacks. However, disclosing of organisation's threat information may increase the risks for the organisation. This process entails risks in various aspects, such as privacy, technical, legal, business, reputation, and organisational aspects. These risks can be evaluated and assessed by providing the right risk model. Cyber threat intelligence enables organisations to continuously monitor and support their business and strategic goals by providing insights regarding existing threat actors and perpetrators trying to target their business. However, sharing such information should be evaluated and assessed to enhance and stimulate cyber threat intelligence sharing, while mitigating the potential adverse effects. Besides, sharing cyber threat intelligence among industry members and governments poses a legal challenge. Thus, it is necessary to provide a model that can help organisations to share cyber threat intelligence and stay compliant with the law.

This chapter presents a thorough discussion of the conclusions of our research, restates the contribution, and identifies issues and opportunities for future research.

6.1 Revisiting the Contribution

The research described in this thesis is novel in that it combines and extends concepts found in risk identification, risk assessment and legal aspects, with the context of cyber threat intelligence within the operations of critical infrastructures.

1. It provides a comprehensive analysis of a cyber incident model to identify the cybersecurity and privacy related threats of disclosing sensitive data and identifying information. It turns out that disclosing cyber incident information consists of risks of disclosing personal information, business information, financial information, and cybersecurity information. The thesis has extended CNIL privacy risk management to cover cybersecurity risks in addition to the privacy risks. Based on this, we calculated the severity of the identified threats associated with each property in both privacy and cybersecurity dimensions. Finally, using these results, this thesis has included a guideline to assist cyber threat intelligence managers to use the STIX incident model while mitigating the risks of sharing (objective 1).

2. This research provides a means to apply risk assessment to the cyber threat intelligence sharing process. It presents a methodology for evaluating the risks of sharing threat intelligence based on quantitive assessments of the properties in the dataset before sharing. It extends the first contribution, so that after it identifies the potential threats associated with sharing a CTI dataset and compute the severity for each property, it proposes an estimation of the likelihood of the threats in case of property disclosure. Finally, it computes the total risk score of sharing a CTI dataset. Based on the risk value, the organisations can select appropriate privacy-preserving techniques to mitigate sharing risk. During the creation of the risk model, the methodology was tested on an open-source dataset and multiple use cases. Then, it empirically evaluated the risk model by using experts' opinion. Three teams

of 24 cybersecurity and privacy experts in total evaluated three different use cases. The results indicate that the experts' selection broadly matches the outcomes produced using our model (objective 2 and objective 4).

3. This research supports the effort to progress cyber threat intelligence sharing by presenting an approach that takes into account the legal dimension. It has suggested adequate techniques for protecting the privacy of data subjects in relation to cyber threat intelligence datasets under the GDPR. Then it defines a policy space as a set of assertions. These assertions consist of the legal grounds under which personal data can be processed under the GDPR. Finally, it builds the model as a decision graph based on the identified assertions, and the final decision will be assigning a specific tag to encode the right level of handling and sharing the cyber threat intelligence dataset (objective 3 and objective 4).

6.2 Future Work

This dissertation identifies and highlights several opportunities and open issues for future study and investigation. The proposed approaches to solving existing challenges in sharing cyber threat intelligence must be further examined. In this section, we conclude by emphasising some open problems and items for future work:

- Chapter 3 presents a detailed analysis of cybersecurity and privacy risks from a perspective of confidentiality. Further investigations are needed to assess other risk types such as business risks, cyber supply chain risks and risks related to integrity threats on cyber intelligence.

- In Chapter 4, the STIX incident dataset's automated analysis is entirely based on the presence or absence of specific properties. We only analyse the content to see whether the property is essentially absent. Further studies are required to refine this analysis using natural language processing techniques to assess sharing CTI dataset risks and build the right sharing decision.

- Chapter 4 proposes a quantitative risk model to assess the aggregate cybersecurity risk of sharing cyberthreat intelligence. The privacy literature indicates that quantification of privacy risk using a simple scale is difficult and often fundamentally inadequate [233][159]. Methods for quantifying privacy risk along with the cybersecurity risk in the context of cyber threat intelligence remain to be explored further.

- Chapter 5 proposed a model to assess the legal requirements for supporting decision-making when sharing cyber threat information. Future research is required to extend our model to assess other privacy and cyber laws.

- There are sophisticated methods of sharing which use privacy-preserving techniques to reduce exposure risks [165]. 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 on the data. Ultimately, the sharing choices need to be balanced, preserving confidentiality with preserving the utility of the analysis. Thus, in the future, it will be important to explore such analysis operations. Besides, a future study investigating the tradeoff between the privacy preservation and utility of processing CTI datasets would be very interesting.

- Organisations have different cyber risk profiles [234] based on sectors, operation standards, needs and regulations. Therefore it is unlikely that a single approach for sharing cyber threat information fits all organisations and governments. For example, there are vast numbers of cyber attacks against the banking sector. Therefore, information sharing platforms and methodologies should be designed to consider sector specific requirements.

- In Chapter 4 and 5, we have included the level of trust as a factor in our evaluation. In the future, it will be important to explore the potential of assessing the trust level among the organisations inside our models. Trust is an essential factor since cyber threat intelligence contains sensitive information. Also, it is vital to avoid potential threat actors getting insight into organisations' analysis and mitigations. Thus a closed and trusted group can provide comprehensive and oriented sharing more than public or untrusted sharing (e.g., Cyber Defence Alliance (CDA) [235] is a group of banks with Anomali -IT security company- [236]).

- It might be beneficial to implement the model to be included and integrated into existing cyber threat intelligence platforms such as MISP [104]. Also, it would be interesting to make our risk assessment model more dynamic by covering all cyber threat intelligence standards such as IOEDF [13]. Furthermore, future work involves a further assessment to confirm our risk assessment model’s practicality by applying it to more real-world and larger scenarios.

- Another possible area of future research would be to investigate how to integrate the proposed risk assessment model in Chapters 3 and 4, and the legal assessment model in Chapter 5 into the organisation’s risk management framework and risk portfolio.

Bibliography

[1] Microsoft Corporation. Differential Privacy for Everyone. *http://download.microsoft.com/download/D/1/F/D1F0DFF5-8BA9-4BDF-8924-7816932F6825/Differential_Privacy_for_Everyone.pdf*, pages 1–6, 2012.

[2] Aharon Chernin Sean Barnum, Desiree Beck and Rich Piazza. STIX Version 1.2.1. Part 1: Overview. *Mitre Corporation*, 2016.

[3] Kaitlin R Boeckl and Naomi B Lefkovitz. NIST Privacy Framework: A Tool for Improving Privacy Through Enterprise Risk Management, Version 1.0. 2020.

[4] T Fredriksson et al. Information economy report, digitalisation, trade and development. *Geneva: UNCTAD, https://unctad.org/system/files/official-document/ier2017_en.pdf*, 2017.

[5] UNCTAD. Digital economy report 2019: Value creation and capture–implications for developing countries. *Naciones Unidas Ginebra, https://unctad.org/system/files/official-document/der2019_en.pdf*, 2019.

[6] TheGuardian. What you need to know about the biggest hack of the US government in years. *https://www.theguardian.com/technology/2020/dec/15/orion-hack-solar-winds-explained-us-treasury-commerce-department*, 2020.

[7] Solarwinds. Solarwinds Orion. *https://www.solarwinds.com/solutions/orion*, 1999.

[8] Ashar Aziz. FireEye - Cybersecurity company. *https://www.fireeye.com*, 2004.

[9] Hispasec Sistemas. VirusTotal. *https://www.virustotal.com/gui/*, 2004.

[10] CISCO. TalosIntelligence. *https://talosintelligence.com*, 2020.

[11] Henry Dalziel. *How to define and build an effective cyber threat intelligence capability*. Syngress, 2014.

[12] MITRE. Structured threat information expression. *http://STIXproject.github.io*, 2014.

[13] R Danyliw, J Meijer, and Y Demchenko. RFC5070-The Incident Object Description Exchange Format. *Internet Engineering Task Force (IETF)*, 5070, 2007.

[14] L. Obrst, P. Chase, and R. Markeloff. Developing an ontology of the cyber security domain. In *Semantic Technology for Intelligence, Defense, and Security, STIDS*, pages 49–56, 2012.

[15] Juniper. Cyber Crime Costs Projected To Reach $2 Trillion by 2019. *https://www.juniperresearch.com/press/press-releases/cybercrime-cost-businesses-over-2trillion-by-2019*, 2019.

[16] Sachin Katti, Balachander Krishnamurthy, and Dina Katabi. Collaborating against Common Enemies. In *Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement*, IMC '05, page 34, USA, 2005. USENIX Association.

[17] Marie Vasek, Matthew Weeden, and Tyler Moore. Measuring the Impact of Sharing Abuse Data with Web Hosting Providers. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, WISCS '16, page 71–80, New York, NY, USA, 2016. Association for Computing Machinery.

[18] Pierangela Samarati and Latanya Sweeney. Generalizing data to provide anonymity when disclosing information. In *PODS*, volume 98, pages 275487–275508, 1998.

[19] Pierangela Samarati. Protecting respondents' identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027, 2001.

[20] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. Calibrating noise to sensitivity in private data analysis. In Shai Halevi and Tal Rabin, editors, *Theory of Cryptography*, pages 265–284, Berlin, Heidelberg, 2006. Springer.

[21] Sabyasachi Mitra and Sam Ransbotham. Information Disclosure and the Diffusion of Information Security Attacks. *Information Systems Research*, 26(3):565–584, September 2015.

[22] Cambridge. Cambridge international dictionary of English. *Cambridge: Cambridge University Press*, 1995.

[23] Samuel D Warren and Louis D Brandeis. The right to privacy. *Harvard Law Review*, pages 193–220, 1890.

[24] Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

[25] Daniel J Solove. A taxonomy of privacy. *U. Pa. L. Rev.*, 154:477, 2005.

[26] Alissa Cooper, Hannes Tschofenig, Bernard Aboba, Jon Peterson, J Morris, Marit Hansen, and Rhys Smith. Privacy considerations for internet protocols. *Internet Architecture Board*, 2013.

[27] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[28] ISO. ISO 25237:2017 Health informatics — Pseudonymization. *https://www.iso.org/standard/63553.html*, 2017.

[29] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3-4):211–407, 2014.

[30] Irit Dinur and Kobbi Nissim. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pages 202–210, 2003.

[31] Michael Barbaro, Tom Zeller, and Saul Hansell. A face is exposed for aol searcher no. 4417749. *New York Times*, 9(2008):8, 2006.

[32] Katie Hafner. if you liked the movie, a Netflix contest may reward you handsomely. *New York Times, https://www.nytimes.com/2006/10/02/technology/02netflix.html?auth=linked-google*, 2, 2006.

[33] Arvind Narayanan and Vitaly Shmatikov. Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (SP 2008)*, pages 111–125. IEEE, 2008.

[34] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity.

ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1):3–es, 2007.

[35] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd International Conference on Data Engineering*, pages 106–115. IEEE, 2007.

[36] Adam J Slagell, Kiran Lakkaraju, and Katherine Luo. FLAIM: A Multi-level Anonymization Framework for Computer and Network Logs. In *Large Installation System Administration Conference(LISA)*, volume 6, pages 3–8, 2006.

[37] Adam Slagell and William Yurcik. Sharing computer network logs for security and privacy: A motivation for new methodologies of anonymization. In *Workshop of the 1st International Conference on Security and Privacy for Emerging Areas in Communication Networks, 2005*, pages 80–89. IEEE, 2005.

[38] Charu C Aggarwal and S Yu Philip. A general survey of privacy-preserving data mining models and algorithms. In *Privacy-preserving data mining, vol 34.*, pages 11–52. Springer, 2008.

[39] Vijay S Iyengar. Transforming data to satisfy privacy constraints. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 279–288, 2002.

[40] Roberto J Bayardo and Rakesh Agrawal. Data privacy through optimal k-anonymization. In *21st International conference on data engineering (ICDE'05)*, pages 217–228. IEEE, 2005.

[41] Jian Xu, Wei Wang, Jian Pei, Xiaoyuan Wang, Baile Shi, and Ada Wai-Chee Fu. Utility-based anonymization using local recoding. In *Proceedings of the*

12th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 785–790, 2006.

[42] Thomas D Wagner, Esther Palomar, Khaled Mahbub, and Ali E Abdallah. Towards an anonymity supported platform for shared cyber threat intelligence. In *International Conference on Risks and Security of Internet and Systems*, pages 175–183. Springer, 2017.

[43] Inc. FIRST.ORG. Traffic Light Protocol (TLP). *https://www.first.org/tlp*, 2001.

[44] Rakesh Agrawal and Ramakrishnan Srikant. Privacy-preserving data mining. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 439–450, 2000.

[45] Vladimir Estivill-Castro and Ljiljana Brankovic. Data swapping: Balancing privacy against precision in mining for logic rules. In *International Conference on Data Warehousing and Knowledge Discovery*, pages 389–398. Springer, 1999.

[46] Sajal Kanti Das, Shrikant Kumar Gupta, and Mohammad Kauser. Micro aggregation Through DBSCAN for PPDM: Privacy-Preserving Data Mining. *International Journal of Advance Research in Science and Engineering-IJARSE*, 1(2):15–21, 2011.

[47] C Dwork. Theory and Applications of Models of Computation. 2008. volume 4978. Springer.

[48] Frank D McSherry. Privacy integrated queries: an extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of data*, pages 19–30, 2009.

[49] Craig Gentry. Fully homomorphic encryption using ideal lattices. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pages 169–178, 2009.

[50] Steven Y Ko, Kyungho Jeon, and Ramsés Morales. The hybrex model for confidentiality and privacy in cloud computing. *HotCloud*, 11:8–8, 2011.

[51] Liam Morris. Analysis of partially and fully homomorphic encryption. *Rochester Institute of Technology*, 10:1–5, 2013.

[52] Michael Naehrig, Kristin Lauter, and Vinod Vaikuntanathan. Can homomorphic encryption be practical? In *Proceedings of the 3rd ACM workshop on Cloud computing security workshop*, pages 113–124, 2011.

[53] Frederik Armknecht and Thorsten Strufe. An efficient distributed privacy-preserving recommendation system. In *The 10th IFIP Annual Mediterranean Ad Hoc Networking Workshop*, pages 65–70. IEEE, 2011.

[54] Adi Shamir. How to share a secret. *Communications of the ACM*, 22(11):612–613, 1979.

[55] Atsushi Waseda and Masakazu Soshi. Consideration for multi-threshold multi-secret sharing schemes. In *2012 International Symposium on Information Theory and its Applications*, pages 265–269. IEEE, 2012.

[56] Ali Nakhaei Amroudi, Ali Zaghain, and Mahdi Sajadieh. A verifiable (k, n, m)-threshold multi-secret sharing scheme based on NTRU cryptosystem. *Wireless Personal Communications*, 96(1):1393–1405, 2017.

[57] David Chaum, Claude Crépeau, and Ivan Damgard. Multiparty unconditionally secure protocols. In *Proceedings of the twentieth annual ACM symposium on Theory of computing*, pages 11–19, 1988.

[58] Peter Bogetoft, Dan Lund Christensen, Ivan Damgard, Martin Geisler, Thomas Jakobsen, Mikkel Krøigaard, Janus Dam Nielsen, Jesper Buus Nielsen, Kurt Nielsen, Jakob Pagter, et al. Secure multiparty computation goes live. In *International Conference on Financial Cryptography and Data Security*, pages 325–343. Springer, 2009.

[59] Dan Bogdanov, Riivo Talviste, and Jan Willemson. Deploying secure multi-party computation for financial data analysis. In *International Conference on Financial Cryptography and Data Security*, pages 57–64. Springer, 2012.

[60] Abid Mehmood, Iynkaran Natgunanathan, Yong Xiang, Guang Hua, and Song Guo. Protection of big data privacy. *IEEE Access*, 4:1821–1834, 2016.

[61] Lei Xu, Chunxiao Jiang, Jian Wang, Jian Yuan, and Yong Ren. Information security in big data: privacy and data mining. *IEEE Access*, 2:1149–1176, 2014.

[62] Vipul Goyal, Omkant Pandey, Amit Sahai, and Brent Waters. Attribute-based encryption for fine-grained access control of encrypted data. In *Proceedings of the 13th ACM conference on Computer and communications security*, pages 89–98, 2006.

[63] John Bethencourt, Amit Sahai, and Brent Waters. Ciphertext-policy attribute-based encryption. In *2007 IEEE symposium on security and privacy (SP'07)*, pages 321–334. IEEE, 2007.

[64] Xavier Boyen and Brent Waters. Anonymous hierarchical identity-based encryption (without random oracles). In *Advances in Cryptology - CRYPTO 2006*, pages 290–307, Berlin, Heidelberg, 2006. Springer Berlin Heidelberg.

[65] Hongbing Cheng, Chunming Rong, Kai Hwang, Weihong Wang, and Yanyan Li. Secure big data storage and sharing scheme for cloud tenants. *China Communications*, 12(6):106–115, 2015.

[66] Lea Kissner and Dawn Song. Privacy-preserving set operations. In *Advances in Cryptology – CRYPTO 2005*, pages 241–257, Berlin, Heidelberg, 2005. Springer Berlin Heidelberg.

[67] Siani Pearson and Marco Casassa-Mont. Sticky policies: An approach for managing privacy across multiple parties. *Computer*, 44(9):60–68, 2011.

[68] D. Miorandi, A. Rizzardi, S. Sicari, and A. Coen-Porisini. Sticky policies: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 32(12):2481–2499, 2020.

[69] Fabio Martinelli, Andrea Saracino, and Mina Sheikhalishahi. Modeling privacy aware information sharing systems: A formal and general approach. In *2016 IEEE Trustcom/BigDataSE/ISPA*, pages 767–774. IEEE, 2016.

[70] Oleksii Osliak, Andrea Saracino, and Fabio Martinelli. A scheme for the sticky policy representation supporting secure cyber-threat intelligence analysis and sharing. *Information and Computer Security*, 27(5):687–710, 2019.

[71] PrivacyPatterns. Privacy patterns. *https://privacypatterns.org*, 2006.

[72] EU FP7 Project PRIPARE. privacypatterns.eu - collecting patterns for better privacy. *https://privacypatterns.eu*, 2016.

[73] Christopher Alexander. *A pattern language: towns, buildings, construction.* Oxford University Press, 1977.

[74] Lawrence A Gordon, Martin P Loeb, William Lucyshyn, and Lei Zhou. The impact of information sharing on cybersecurity underinvestment: A real options perspective. *Journal of Accounting and Public Policy*, 34(5):509–519, 2015.

[75] Chenfeng Vincent Zhou, Christopher Leckie, and Shanika Karunasekera. A survey of coordinated attacks and collaborative intrusion detection. *Computers and Security*, 29(1):124–140, 2010.

[76] Alexandre Dulaunoy, Gérard Wagener, Marc Stiefer, and Cynthia Wagner. The void–an interesting place for network security monitoring. In *Proceedings of the 30th TERENA networking conference (TNC'14). Dublin, Ireland.* Citeseer, 2014.

[77] Sami Mokaddem, Gérard Wagener, and Alexandre Dulaunoy. AIL-The design and implementation of an Analysis Information Leak framework. In *2018 IEEE International Conference on Big Data (Big Data)*, pages 5049–5057. IEEE, 2018.

[78] Novetta Threat Research Group et al. Operation Blockbuster: Unraveling the Long Thread of the Sony Attack. *https://operationblockbuster.com, February*, 24, 2016.

[79] Aaron Boyd. How FBI Cyber Division helps agencies investigate intrusions. *https://www.federaltimes.com/enterprise-view/2015/10/30/how-fbi-cyber-division-helps-agencies-investigate-intrusions*, 2015.

[80] US-CERT. Targeted Destructive Malware- Alert (TA14-353A). *https://www.us-cert.gov/ncas/alerts/TA14-353A*, 2014.

[81] Tom Bergin and Jim Finkle. Exclusive: Swift confirms new cyber thefts, hacking tactics. *https://www.reuters.com/article/uk-usa-cyber-swift-idUKKBN1412NU?edition-redirect=uk, December*, 12, 2016.

[82] Sergei Shevchenko and Adrian Nish. Cyber heist attribution. *BAE Systems Threat Research Blog, https://www.baesystems.com/en/cybersecurity/cyber-heist-attribution*, 2016.

[83] Tom Bergin and Nathan Layne. Special report: Cyber thieves exploit banks' faith in swift transfer network. *https://es.reuters.com/article/us-cyber-heist-swift-specialreport/special-report-cyber-thieves-exploit-banks-faith-in-swift-transfer-network-idUSKCN0YB0DD*, 2016.

[84] Rob McMillan. Definition: threat intelligence. *https://www.gartner.com/en/documents/2487216/definition-threat-intelligence*, 2013.

[85] S Caltagirone. Threat intelligence definition: What is old is new again. *http://www.activeresponse.org/threat-intelligence-definition-old-new*, 2016.

[86] Robert M Lee. Intelligence defined and its impact on cyber threat intelligence. *https://www.robertmlee.org/intelligence-defined-and-its-impact-on-cyber-threat-intelligence*, 2, 2016.

[87] Chairman of the Joint Chiefs of Staff. Joint publication 2-0: Joint intelligence Committee. *https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp2_0.pdf*, 2013.

[88] LB Metcalf, Eric Hatleback, and Jonathan M Spring. Blacklist ecosystem analysis: 2016 update. *Software Engineering Institute, CERT Coordination Center, Pittsburgh, PA*, 2016.

[89] David Chismon and Martyn Ruks. Threat intelligence: Collecting, analysing, evaluating. *MWR InfoSecurity Ltd, https://www.foo.be/docs/informations-sharing/Threat-Intelligence-Whitepaper.pdf*, 2015.

[90] Maximiliano E Korstanje. *Threat mitigation and detection of cyber warfare and terrorism activities*. IGI Global, 2016.

[91] MITRE. Apt29. *https://attack.mitre.org/groups/G0016*, 2017.

[92] Christopher Johnson, Mark Badger, David Waltermire, Julie Snyder, and Clem Skorupka. Guide to cyber threat information sharing. Technical report, https://www.nist.gov/publications/guide-cyber-threat-information-sharing, National Institute of Standards and Technology, 2016.

[93] NCSC. The national cyber security centre. *https://www.ncsc.gov.uk*, 2016.

[94] UK CERT. Cyber-security Information Sharing Partnership (CiSP). *https://www.ncsc.gov.uk/section/keep-up-to-date/cisp*, 2015.

[95] Scott J Roberts and Rebekah Brown. *Intelligence-Driven Incident Response: Outwitting the Adversary.* O'Reilly Media, Inc., 2017.

[96] Frank Fransen and Richard Kerkdijk. Cyber threat intelligence sharing through national and sector-oriented communities. In *Collaborative Cyber Threat Intelligence*, pages 187–224. Auerbach Publications, 2017.

[97] Greg Farnham and Kees Leune. Tools and standards for cyber threat intelligence projects. *https://www.sans.org/reading-room/whitepapers/warfare/tools-standards-cyber-threat-intelligence-projects-34375, SANS Institute*, 2013.

[98] Abuse.Ch. Ransomware tracker. *https://abuse.ch*, 2016.

[99] blocklist.de. Fail2ban reporting service. *http://www.blocklist.de/en/index.html*, 2016.

[100] CIRCL. Malware Information Sharing Platform MISP—A Threat Sharing Platform. *https://www.circl.lu/services/misp-malware-information-sharing-platform*, 2018.

[101] C. Sauerwein, C. Sillaber, Andrea Mussmann, and R. Breu. Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives. *Business informatics and applied informatics*, 2017.

[102] CIF. Collective intelligence framework (cif). *http://csirtgadgets.org/*, 2012.

[103] D Andre, A Dereszowski, A Dulaunoy, A Iklody, C Vandeplas, and R Vinot. MISP: Malware Information Sharing Platform. *http://www.misp-project.org*, 2011.

[104] MISP. MISP Communities. *https://www.misp-project.org/communities*, 2016.

[105] Yeti. YETI Platform. *https://yeti-platform.github.io*, 2017.

[106] Mike Goffin. CRITs: Collaborative Research Into Threats. *The MITRE Corporation, http://crits.github.io*, 2014.

[107] Palo Alto. Minemeld. *https://github.com/PaloAltoNetworks/minemeld*, 2016.

[108] Anomali. Anomali threatstream. *https://www.anomali.com/*, 2017.

[109] EclecticIQ. Eclecticiq platform. *https://www.eclecticiq.com/platform*, 2014.

[110] LookingGlass. LookingGlass Manage Intelligence. *https://www.lookingglasscyber.com/products/threat-platforms/scoutthreat*, 2015.

[111] NC4. NC4 Soltra Edge. *https://www.celerium.com*, 2014.

[112] Micro Focus. Micro focus threat central. *https://www.microfocus.com/en-us/solutions/application-security*, 2015.

[113] ThreatConnect. Threatconnect. *https://threatconnect.com*, 2013.

[114] ThreatQuotient. Threatquotient threatq. *https://www.threatq.com/threat-intelligence-platform*, 2015.

[115] TruSTAR. Trustar technology. *https://www.trustar.co*, 2014.

[116] Philip Anderson. Tools and Methodologies to Support Co-operation between CSIRTs and Law Enforcement. *ENISA, https://www.enisa.europa.eu/publications/tools-and-methodologies-to-support-cooperation-between-csirts-and-law-enforcement*, 2018.

[117] Christian Sillaber, Clemens Sauerwein, Andrea Mussmann, and Ruth Breu. Data quality challenges and future research directions in threat intelligence sharing practice. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, pages 65–70, 2016.

[118] Sarah Brown, Joep Gommers, and Oscar Serrano. From cyber security information sharing to threat management. In *Proceedings of the 2nd ACM workshop on information sharing and collaborative security*, pages 43–49, 2015.

[119] MITRE. Cyber observable expression. *http://cyboxproject.github.io*, 2011.

[120] Sean Barnum. Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX). *Mitre Corporation*, 11:1–22, 2012.

[121] Julie Connolly, Mark Davidson, and Charles Schmidt. Trusted Automated eXchange of Indicator Information (TAXII). *The MITRE Corporation*, pages 1–20, 2014.

[122] The MITRE Corporation. STIX Schemas Test. *https://github.com/STIXProject/schemas-test*, 2015.

[123] Desiree Beck, Ivan Kirillov, and Rich Piazza. STIX ™ Version 2.0. Part 2: STIX Objects. *Mitre Corporation*, (June):1–58, 2017.

[124] Praetox-Technologies. LOIC (Low Orbit Ion Cannon). *https://sourceforge.net/projects/loic*, 2014.

[125] R Danyliw. The Incident Object Description Exchange Format Version 2. *https://www.rfc-editor.org/info/rfc5070*, 2016.

[126] VERIZON. The Vocabulary for Event Recording and Incident Sharing (VERIS). *http://veriscommunity.net/*, 2016.

[127] The MITRE Corporation. About STIX. *https://STIXproject.github.io/about/*, 2018.

[128] YaraRules Project. YaraRules Project. *https://yararules.com*, 2013.

[129] SIGMA. SIGMA - Generic Signature Format for SIEM Systems. *https://github.com/Neo23x0/sigma*, 2016.

[130] Florian Menges and Günther Pernul. A comparative analysis of incident reporting formats. *Computers and Security*, 73:87 – 101, 2018.

[131] Ross Anderson and Tyler Moore. The Economics of Information Security. *American Association for the Advancement of Science*, 314(5799):610–613, 2006.

[132] Boris Petrenj, Emanuele Lettieri, and Paolo Trucco. Information sharing and collaboration for critical infrastructure resilience–a comprehensive review on barriers and emerging capabilities. *International journal of critical infrastructures*, 9(4):304–329, 2013.

[133] Eric Luiijf and Marieke Klaver. On the sharing of cyber security information. In Mason Rice and Sujeet Shenoi, editors, *Critical Infrastructure Protection IX*, pages 29–46, Cham, 2015. Springer International Publishing.

[134] Tomas Sander and Joshua Hailpern. Ux aspects of threat information sharing platforms: An examination & lessons learned using personas. In *Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security*, pages 51–59, 2015.

[135] David Mann, J Brooks, and Joe DeRosa. The Relationship between Human and Machine-Oriented Standards and the Impact to Enterprise Systems Engineering. *The MITRE Corporation, Bedford, MA, https://www.mitre.org/sites/default/files/pdf/10*$_2$*335.pdf*, 2011.

[136] Ryan Heartfield, George Loukas, Sanja Budimir, Anatolij Bezemskij, Johnny RJ Fontaine, Avgoustinos Filippoupolitis, and Etienne Roesch. A taxonomy of cyber-physical threats and impact in the smart home. *Computers and Security*, 78:398–428, 2018.

[137] Ioannis Agrafiotis, Jason R C Nurse, Michael Goldsmith, Sadie Creese, and David Upton. A taxonomy of cyber-harms: Defining the impacts of cyber-attacks and understanding how they propagate. *Journal of Cybersecurity*, 4(1), 10 2018.

[138] James Tarala and Kelli K. Tarala. Open Threat Taxonomy. *https://www.auditscripts.com/resources/open*$_t$*hreat*$_t$*axonomy*$_v$*1.1a.pdf*, 2015.

[139] MITRE. Lazarus Group. *https://attack.mitre.org/groups/G0032*, 2017.

[140] James L Cebula and Lisa R Young. A taxonomy of operational cyber security risks. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 2010.

[141] James J Cebula, Mary E Popeck, and Lisa R Young. A taxonomy of operational cyber security risks version 2. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 2014.

[142] RM Blank, PD Gallagher, Joint Task Force Transformation Initiative Interagency Working Group, et al. NIST Special Publication 800-53, Security and Privacy Controls for Federal Information Systems and Organizations, Revision 4. *Washington, DC: National Institute of Standards and Technology (NIST)*, 2013.

[143] Office of the Law Revision Counsel. Federal information systems management act of 2002. *https://www.govinfo.gov/content/pkg/PLAW-107publ347/pdf/PLAW-107publ347.pdf*, 2002.

[144] Steven M. Launius. Evaluation of Comprehensive Taxonomies for Information Technology Threats. *https://www.sans.org/reading-room/whitepapers/threatintelligence/evaluation-comprehensive-taxonomies-information-technology-threats-38360*, 2018.

[145] Louis Marinos. ENISA Threat Taxonomy: A tool for structuring threat information. *ENISA, Heraklion*, 2016.

[146] Eric W Burger, Michael D Goodman, Panos Kampanakis, and Kevin A Zhu. Taxonomy model for cyber threat intelligence information exchange technologies. In *Proceedings of the 2014 ACM Workshop on Information Sharing & Collaborative Security*, pages 51–60, 2014.

[147] Barack Obama. Executive order 13636: Improving critical infrastructure cybersecurity. *Federal Register*, 78(33):11739, 2013.

[148] M Eric Johnson. Information risk of inadvertent disclosure: An analysis of file-sharing risk in the financial supply chain. *Journal of Management Information Systems*, 25(2):97–124, 2008.

[149] European Union. Regulation 2016/679 of the European parliament and the Council of the European Union of 27 April 2016 on the protection of

natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/. *Official Journal of the European Union (OJ)*, 59(May):1–88, 2016.

[150] Laurence Kalman. The GDPR and NIS Directive A new age of accountability, security and trust. *https://owasp.org/www-chapter-cambridge*, 2017.

[151] ENISA. European Union Agency for Network and Information Security. *https://www.enisa.europa.eu*, 2004.

[152] ENISA. European union agency for network and information security. *URL: https://www. enisa. europa. eu/topics/cyber-exercises/cyber-europe-programme (accessed 01/12/2020)*, 2004.

[153] Department for Digital Culture Media Sport The Rt Hon Oliver Dowden CBE MP and Matt Warman MP. Post-Implementation Review of the Network and Information Systems Regulations 2018. CP 242(May), 2020.

[154] Matthew Field. WannaCry cyber attack cost the NHS £92m as 19,000 appointments cancelled. *the Telegraph, https://www.telegraph.co.uk/technology/2018/10/11/wannacry-cyber-attack-cost-nhs-92m-19000-appointments-cancelled*, 11, 2018.

[155] European Union. Regulation (EU) 2019/881 of the European Parliament and of the Council of 17 April 2019 on ENISA (the European Union Agency for Cybersecurity) and on information and communications technology cybersecurity certification and repealing Regulation (EU) No 526/2013 (Cybersecurity Act)(Text with EEA relevance). *Official Journal of the European Union, L*, 151:15, 2019.

[156] Shirley Radack. Managing information security risk: organization, mission and information system view. Technical report, National Institute of Standards and Technology, 2011.

[157] Ronald S Ross. Guide for conducting risk assessments (NIST sp-800-30rev1). *The National Institute of Standards and Technology (NIST), Gaithersburg*, 2012.

[158] Christopher Alberts, Audrey Dorofee, James Stevens, and Carol Woody. Introduction to the OCTAVE Approach. Technical report, Carnegie-Mellon Univ Pittsburgh Pa Software Engineering Inst, 2003.

[159] Isabel Wagner and Eerke Boiten. Privacy Risk Assessment: From Art to Science, by Metrics. *Data Privacy Management, Cryptocurrencies and Blockchain Technology*, page 225–241, 2018.

[160] Michael Howard and Steve Lipner. The Security Development Lifecycle: SDL: A Process for Developing Demonstrably More Secure Software. *Microsoft Press*, page 352, 2006.

[161] Kim Wuyts and Wouter Joosen. LINDDUN privacy threat modeling: a tutorial. *CW Reports*, 2015.

[162] Jaspreet Bhatia, Travis D Breaux, Liora Friedberg, Hanan Hibshi, and Daniel Smullen. Privacy risk in cybersecurity data sharing. In *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security*, pages 57–64, 2016.

[163] Gina Fisk, Calvin Ardi, Neale Pickett, John Heidemann, Mike Fisk, and Christos Papadopoulos. Privacy principles for sharing cyber security data. In *2015 IEEE Security and Privacy Workshops*, pages 193–197. IEEE, 2015.

[164] George D Webster, Ryan L Harris, Zachary D Hanif, Bruce A Hembree, Jens Grossklags, and Claudia Eckert. Sharing is caring: Collaborative analysis and real-time enquiry for security analytics. In *2018 IEEE International Conference on Internet of Things (iThings) and (GreenCom) and (CPSCom) and (SmartData)*, pages 1402–1409. IEEE, 2018.

[165] José M de Fuentes, Lorena González-Manzano, Juan Tapiador, and Pedro Peris-Lopez. Pracis: Privacy-preserving and aggregatable cybersecurity information sharing. *Computers and Security*, 69:127–141, 2017.

[166] Riyana Lewis, Panos Louvieris, Pamela Abbott, Natalie Clewley, and Kevin Jones. Cybersecurity information sharing: a framework for information security management in UK SME supply chains. 2014.

[167] Jinsoo Shin, Hanseong Son, and Gyunyoung Heo. Cyber security risk evaluation of a nuclear I&C using BN and ET. *Nuclear Engineering and Technology*, 49(3):517–524, 2017.

[168] M Ugur Aksu, M Hadi Dilek, E İslam Tatlı, Kemal Bicakci, H Ibrahim Dirik, M Umut Demirezen, and Tayfun Aykır. A quantitative CVSS-based cyber security risk assessment methodology for IT systems. In *2017 International Carnahan Conference on Security Technology (ICCST)*, pages 1–8. IEEE, 2017.

[169] Tawei Wang, Karthik N Kannan, and Jackie Rees Ulmer. The association between the disclosure and the realization of information security risk factors. *Information Systems Research*, 24(2):201–218, 2013.

[170] Valentina Viduto, Carsten Maple, Wei Huang, and David LóPez-PeréZ. A novel risk assessment and optimisation model for a multi-objective network security countermeasure selection problem. *Decision Support Systems*, 53(3):599–610, 2012.

[171] Daniel M Best, Jaspreet Bhatia, Elena S Peterson, and Travis D Breaux. Improved cyber threat indicator sharing by scoring privacy risk. In *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, pages 1–5. IEEE, 2017.

[172] Kaniz Fatema, David W Chadwick, and Brendan Van Alsenoy. Extracting access control and conflict resolution policies from european data protection law. In *IFIP PrimeLife International Summer School on Privacy and Identity Management for Life*, pages 59–72. Springer, 2011.

[173] EU Parliament. Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal of the European Communities, number L*, 281:31–50, 1995.

[174] Peter Doorn and Emily Thomas. Tagging Privacy-Sensitive Data According to the New European Privacy Legislation: GDPR DataTags - a Prototype. *https://dans.knaw.nl/en/current/first-gdpr-datatags-results-presented-in-workshop*, 2017.

[175] Heiko Tjalsma. DANS Data Archiving and Networked Services. *https://easy.dans.knaw.nl*, 2012.

[176] Travis D Breaux and Annie I Antón. A systematic method for acquiring regulatory requirements: A frame-based approach. *RHAS-6, Delhi, India*, 2007.

[177] Travis Breaux and Annie Antón. Analyzing regulatory rules for privacy and security requirements. *IEEE Transactions on Software Engineering*, 34(1):5–20, 2008.

[178] Erich Schweighofer, Vinzenz Heussler, and Peter Kieseberg. Privacy by design data exchange between CSIRTs. In *Privacy Technologies and Policy*, pages 104–119. Springer International Publishing, 2017.

[179] Clare Sullivan and Eric Burger. “in the public interest”: The privacy implications of international business-to-business sharing of cyber-threat intelligence. *Computer Law and Security Review*, 33(1):14–29, 2017.

[180] Adham Albakri, Eerke Boiten, and Rogério De Lemos. Risks of sharing cyber incident information. In *Proceedings of the 13th International Conference on Availability, Reliability and Security*, ARES 2018, Hamburg, Germany, 2018. Association for Computing Machinery.

[181] Ken Naganuma, Masayuki Yoshino, Hisayoshi Sato, and Yoshinori Sato. Privacy-preserving analysis technique for secure, cloud-based big data analytics. *Hitachi Rev*, 63(9):577–583, 2014.

[182] Commission Nationale de l'Informatique et des Libertés. Methodology for Privacy Risk Management; How to implement the Data Protection Act, 2012.

[183] Defense Use Case. Analysis of the cyber attack on the Ukrainian power grid. *Electricity Information Sharing and Analysis Center (E-ISAC)*, 388, 2016.

[184] Blake E Strom, Andy Applebaum, Doug P Miller, Kathryn C Nickels, Adam G Pennington, and Cody B Thomas. MITRE ATT&CK™: Design and Philosophy. *Technical report, https://www.mitre.org/sites/default/files/publications/pr-18-0944-11-mitre-attack-design-and-philosophy.pdf*, 2018.

[185] Ian Ahl. Privileges and Credentials: Phished at the Request of Counsel. *https://www.fireeye.com/blog/threat-research/2017/06/phished-at-the-request-of-counsel.html*, 2017.

[186] NIST. CVE-2016-6662 Detail. *https://nvd.nist.gov/vuln/detail/CVE-2016-6662*, 2016.

[187] NIST. CVE-2014-7169 Detail. *https://nvd.nist.gov/vuln/detail/CVE-2014-7169*, 2014.

[188] Ossmann. Throwing Star LAN Tap. *https://ossmann.blogspot.com/2011/02/throwing-star-lan-tap.html*, 2011.

[189] Hak5. Stealing Files with the USB Rubber Ducky – USB Exfiltration Explained. *https://www.hak5.org/blog/main-blog/stealing-files-with-the-usb-rubber-ducky-usb-exfiltration-explained*, 2005.

[190] Robert Mcmillan. The Pwn Plug is a little white box that can hack your network. *https://arstechnica.com/information-technology/2012/03/the-pwn-plug-is-a-little-white-box-that-can-hack-your-network/*, 2012.

[191] Robert Falcone Bryan Lee, Mike Harbison. Sofacy Attacks Multiple Government Entities. *https://unit42.paloaltonetworks.com/unit42-sofacy-attacks-multiple-government-entities*, 2018.

[192] F-SECURE. THE DUKES 7 years of Russian cyberespionage. *https://www.f-secure.com/documents/996508/1030745/dukes*$_w$*hitepaper.pdf*, 2016.

[193] Tomáš Foltýn. OceanLotus ships new backdoor using old tricks. *https://www.welivesecurity.com/2018/03/13/oceanlotus-ships-new-backdoor*, 2018.

[194] FireEye. APT37 (REAPER) The Overlooked North Korean Actor. *https://www2.fireeye.com/rs/848-DID-242/images/rpt*$_APT$37*.pdf*, 2018.

[195] Edmund Brumaghin, Ross Gibb, Warren Mercer, Matthew Molyett, and Craig Williams. Talos Blog || Cisco Talos Intelligence Group - Comprehensive Threat Intelligence: CCleanup: A Vast Number of Machines at Risk. *https://blog.talosintelligence.com/2017/09/avast-distributes-malware.html*, 2017.

[196] Gavin O'Gorman and Geoff McDonald. *The Elderwood Project*. Symantec Corporation, https://www.infopoint-security.de/medien/the-elderwood-project.pdf, 2012.

[197] Microsoft. Behavior monitoring combined with machine learning spoils a massive Dofoil coin mining campaign. *https://www.microsoft.com/security/blog/2018/03/07/behavior-monitoring-combined-with-machine-learning-spoils-a-massive-dofoil-coin-mining-campaign*, 2018.

[198] Symantec. Buckeye cyberespionage group shifts gaze from US to Hong Kong | Symantec Connect Community. *https://www.symantec.com/connect/blogs/buckeye-cyberespionage-group-shifts-gaze-us-hong-kong*, 2016.

[199] GReAT. The Great Bank Robbery : the Carbanak. *https://securelist.com/the-great-bank-robbery-the-carbanak-apt/68732*, pages 1–12, 2015.

[200] Iverson5. Cobalt Strike DLL. *https://cobaltstrike.com/downloads/csmanual38.pdf*, pages 3–5, 2017.

[201] Stephane M Meystre, F Jeffrey Friedlin, Brett R South, Shuying Shen, and Matthew H Samore. Automatic de-identification of textual documents in the electronic health record: a review of recent research. *BMC medical research methodology*, 10(1):70, 2010.

[202] CESG. Common Cyber Attacks: Reducing The Impact. *https://goo.gl/sdozWr*, page 17, 2015.

[203] MITRE. TTPType. *https://stixproject.github.io/data-model/1.2/ttp/TTPType*, 2016.

[204] U.S. District Court. US Dept. of Justice Indictment Chinese Hack. *https://www.justice.gov/iso/opa/resources/5122014519132358461949.pdf*, 2014.

[205] OASIS. Open standards. Open source. *https://www.oasis-open.org/*, 2012. [Online; accessed 2020-12-10].

[206] OASIS. Comparing STIX 1.X/CybOX 2.X with STIX 2. *https://oasis-open.github.io/cti-documentation/STIX/compare*, 2018.

[207] Adham Albakri, Eerke Boiten, and Richard Smith. Risk Assessment of Sharing Cyber Threat Intelligence. In *European Symposium on Research in Computer Security*, pages 92–113. Springer, 2020.

[208] MITRE. STIX Incident Model. *https://STIXproject.github.io/data-model/1.2/incident/IncidentType*, 2018.

[209] Douglas Hubbard and Dylan Evans. Problems with scoring methods and ordinal scales in risk assessment. *IBM Journal of Research and Development*, 54(3):2–1, 2010.

[210] Rebecca M. Blank and Acting Secretary. Guide for Conducting Risk Assessments, 2011.

[211] MITRE. Ttptype. *https://STIXproject.github.io/data-model/1.2/ttp/TTPType*, 2016. [Online; accessed 2020-12-06].

[212] Ronen Feldman, James Sanger, et al. *The text mining handbook: advanced approaches in analyzing unstructured data.* Cambridge university press, 2007.

[213] Joseph L Fleiss. Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378, 1971.

[214] Scott E Coull, Charles V Wright, Fabian Monrose, Michael P Collins, Michael K Reiter, et al. Playing devil's advocate: Inferring sensitive information from anonymized network traces. In *Ndss*, volume 7, pages 35–47, 2007.

[215] Alexander D Kent and Lorie M Liebrock. Secure communication via shared knowledge and a salted hash in ad-hoc environments. In *2011 IEEE 35th Annual Computer Software and Applications Conference Workshops*, pages 122–127. IEEE, 2011.

[216] Frederik Armknecht, Colin Boyd, Christopher Carr, Kristian Gjøsteen, Angela Jäschke, Christian A Reuter, and Martin Strand. A guide to fully homomorphic encryption. *IACR Cryptol. ePrint Arch.*, 2015:1192, 2015.

[217] Andrew C Yao. Protocols for secure computations. In *23rd annual symposium on foundations of computer science (sfcs 1982)*, pages 160–164. IEEE, 1982.

[218] MITRE. Red October. *https://github.com/STIXProject/schemas-test/blob/master/veris*, 2015. [Online; accessed 2019-12-09].

[219] Kaspersky. Attackers created unique, highly-flexible malware to steal data and geopolitical intelligence from target victims' computer systems, mobile phones and enterprise network equipment. *https://www.kaspersky.com/about/press-releases/2013*$_k$*aspersky − lab − identifies − operation − −red − october − −an − advanced − cyber − espionage − campaign − targeting − diplomatic − and − government − institutions − worldwide*, 2013. [*Online*; *accessed*2020 − 02 − 25].

[220] RITICS. Air4ics: Agile incident response for industrial control systems. *https://ritics.org/projects/air4ics-agile-incident-response-for-industrial-control-systems*, 2019. [Online; accessed 2020-02-17].

[221] Adham Albakri, Eerke Boiten, and Rogério De Lemos. Sharing Cyber Threat Intelligence Under the General Data Protection Regulation. In *Privacy Technologies and Policy*, pages 28–41, Cham, 2019. Springer International Publishing.

[222] Michael Bar-Sinai, Latanya Sweeney, and Merce Crosas. Datatags, data handling policy spaces and the tags language. In *2016 IEEE Security and Privacy Workshops (SPW)*, pages 1–8. IEEE, 2016.

[223] Latanya Sweeney, Mercè Crosas, and Michael Bar-Sinai. Sharing sensitive data with confidence: The datatags system. *Technology Science*, 2015.

[224] ENISA. Considerations on the Traffic Light Protocol. *ENISA, https://www.enisa.europa.eu/topics/csirts-in-europe/glossary/considerations-on-the-traffic-light-protocol*, 2019.

[225] CIRCL. Legal compliance and CSIRT activities. *https://github.com/CIRCL/compliance*, 2018.

[226] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[227] Xiaokui Xiao and Yufei Tao. Personalized privacy preservation. In *Proceedings of the 2006 ACM SIGMOD international conference on Management of data*, pages 229–240, 2006.

[228] Mehmet Ercan Nergiz, Maurizio Atzori, and Chris Clifton. Hiding the presence of individuals from shared databases. In *Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 665–676, 2007.

[229] CIRCL. AIL information leaks analysis and the GDPR in the context of collection, analysis and sharing information leaks. *https://www.circl.lu/assets/files/information-leaks-analysis-and-gdpr.pdf*, 2018.

[230] Sam Thielman. Yahoo hack: 1bn accounts compromised by biggest data breach in history. *The Guardian*, 15:2016, 2016.

[231] MITRE. Sensitive information belonging to jobseekers has been put at risk. *https://github.com/STIXProject/schemas-test/blob/master/veris/0012CC25-9167-40D8-8FE3-3D0DFD8FB6BB.xml*, 2015.

[232] Group-IB Oleg Skulkin. DarkHydrus. *DarkHydrus, Group G0079 | MITRE ATTCK®*, 2018.

[233] Fenia Ferra, Isabel Wagner, Eerke Boiten, Lee Hadlington, Ismini Psychoula, and Richard Snape. Challenges in assessing privacy impact: Tales from the front lines. *Security and Privacy*, 3(2):e101, 2020.

[234] Denise E Zheng and James A Lewis. Cyber threat information sharing. *Center for Strategic and International Studies*, 2015.

[235] Anomali. Cyber Defence Alliance (CDA) partners with Anomali to better enable sharing of Threat Intelligence among banking members. *https://www.anomali.com/news-events/press/cyber-defence-alliance-cda-partners-with-anomali-to-better-enable-sharing-of-threat-intelligence-among-banking-members*, 2020.

[236] Anomali. Anomali - IT security company. *https://www.anomali.com*, 2013.