Horizon 2020 Programme
Instrument: Innovation Action

Proactive Risk Management through Improved Cyber Situational Awareness

Start Date of Project: 2016-09-01

Duration: 36 months

D5.1 Threat Intelligence Sharing: State of the Art and Requirements

Deliverable Details	
Deliverable Number	**D5.1**
Revision Number	**E**
Author(s)	**TUDA/OXF/AIT**
Due Date	**0517**
Delivered Date	**29/05/17**
Reviewed by	**CESNET/RoEduNet/PSNC/TheEmailLaundry/GMV**
Dissemination Level	**PU**
Contact Person EC	**Georgios Kaiafas**

Contributing Partners	
1.	**TheEmailLaundry (reviewer)**
2.	**GMV (reviewer)**
3.	**CESNET (reviewer)**
4.	**RoEduNet (reviewer)**
5.	**PSNC (reviewer)**

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under grant agreement no 700071.

Revision History

Revision	By	Date	Changes
E	AIT	29/05/17	Version Submitted to Agency
A19	AIT, OXF, TUDA	29/05/17	Final revisions and checks.
A18	CESNET, TheEmailLaundry, TUDA	24/05/17	Considered CESNET and EML reviews and added responses
A17	PSNC, TUDA	23/05/17	Considered PSNC review and OXF responses
A16	TUDA	18/05/17	Addition of list of listings, tables; Typos fixed; Added references, Addressed comments (for TUDA). Fixed cross-referencing throughout the document.
A15	OXF	18/05/17	Refs added, minor corrections, comments resolved
A14	TUDA, OXF	17/05/17	New content in 3.2.2, 3.2.4, 3.3.4
A13	TUDA	16/05/17	Revised content in 3.2.3 and 3.3.3
A12	AIT	15/05/17	Revised content in 3.2.1, 3.3, 3.3.1, and 3.3.2
A11	TUDA, AIT, OXF	12/0517	Revised content in Chapter 2 and Chapter 3
A10	TUDA, AIT	19/04/17	Content added in Sec 2.4, 3.4
A9	TUDA	19/04/2017	Content added in Sec. 2.2, 3.4
A8	OXF, AIT, TUDA	12/04/2017	Content added in Chap. 1, Sec. 2.1, 2.2, 2.3, 2.4. Sec. 3.4
A7	TUDA	14/03/2017	Shaped the deliverable and assigned content writing tasks
A6	TUDA, OXF, AIT	10/03/2017	Revised the content: new sections added
A5	OXF, TUDA, AIT	08/02/2017	Revised the content and ToC
A4	TUDA	03/11/2016	Updated with individual comments
A3	OXF	11/01/2016	Updated with individual comments
A2	OXF	27/10/2016	Revised ToC, chapter refinement
A1	TUDA	26/10/2016	Initial ToC

Abbreviations

ACTIC	Arizona Counter Terrorism Intelligence Center
ACTRA	Arizona Cyber Threat Response Alliance
API	Application Programming Interface
BPEL	Business Process Execution Language
CDC	Centers for Disease Control and Prevention
CERT	Computer Emergency Response Team
CISA	Cyber Information Sharing Agreements
COMPAS	Centre on Migration, Policy and Society
CPNI	Centre for the Protection of National Infrastructure
CSA	Cyber Security Awareness
CSIRT	Computer Security Incident Response Team
CSV	Character-Separated Values
CTI	Cyber Threat Intelligence
DIKW	Data, Information, Knowledge, Wisdom
DoS	Denial-Of-Service
DDoS	Distributed Denial-Of-Service
DPA	Data Protection Authority
EC	European Commission
EML	TheEmailLaundry
ENISA	European Union Agency for Network and Information Security
EU	European Union
FIRST	Forum for Incident Response and Security Team
GDPR	General Data Protection Regulations
HASL	Handling, Action, Sharing and Licensing
HCI	Human Computer Interaction
HIPAA	Heath Insurance Portability and Accountability Act
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
ID	IDentification
IDEA	Intrusion Detection Extensible Alert
IDMEF	Intrusion Detection Message Exchange Format
IDS	Intrusion Detection System
IEP	Information Exchange Policy
IOC	Indicators of Compromise
IODEF	Incident Object Description Exchange Format
IP	Internet Protocol
IPS	Intrusion Protection System
ISA	Information Sharing Agreements
JSON	Javascript Object Notation
KE	Knowledge Exchange
LSTM	Long Short Term Memory
MACCSA	Multinational Alliance for Collaborative Cyber Situational Awareness
MISP	Malware Information Sharing Platform
ML	Machine Learning
NAT-PMP	Network Address Translation- Port Mapping Protocol
NATO	North Atlantic Treaty Organisation

NDA	Non-Disclosure Agreement
NIST	National Institute of Standards and Technology
NREN	National Research and Education Network
NTP	Network Time Protocol
RDP	Remote Desktop Protocol
SDK	Software Development Kit
SOA	Service-Oriented Architecture
SME	Small to Mid-size Enterprise
SNMP	Simple Network Management Protocol
SSDP	Simple Service Discovery Protcol
SSH	Secure Shell
STIX	Structured Threat Information eXpression
TAXII	Trusted Automated eXchange of Indicator Information
TDK	Threat and Defence Knowledge
TI	Threat Intelligence
TIP	Threat Intelligence Platform
TLP	Traffic Light Protocol
TM	TeleManagement Forum
UK	United Kingdom
URL	Uniform Resource Locator
XML	eXtensible Markup Language

Executive Summary

This report describes the PROTECTIVE threat intelligence-sharing framework. The document aims firstly to clarify the terminology used with regard to Threat Intelligence (TI). We provide the reader with an in-depth discussion of respective literature. The deliverable also analyses the state of the art in the area of information types, and models, methods and mechanisms for TI sharing. Finally, we also describe the requirements and specifications of TI sharing in the project tool.

The document is structured as follows. First, Section 1 serves as an introduction to the field of TI by discussing the basic terminology that will be utilised in the project as well as the differences of the various terms. In addition, the section specifically discusses the way PROTECTIVE defines TI-related terms. Section 2 gives an overview of the state of the art in the area of TI and particularly on the plethora of concepts and building blocks that exist. Moreover, it analyses the topic of compliance (i.e., the legal aspects of TI sharing) and concludes with case studies (both from potential use cases of PROTECTIVE, and from existing related literature) of practical examples in which compliance is mandatory. Section 3 acts as the core contribution of this deliverable by providing requirements and specifications for TI sharing in the PROTECTIVE ecosystem. In particular, the section describes the architecture of the PROTECTIVE TI Sharing system and its corresponding internal components. Lastly, Section 4 concludes this document.

Horizon 2020 Programme
Instrument: Innovation Action

The research leading to these results has received funding from the European Union's Horizon 2020 Research and Innovation Programme, under grant agreement no 700071.

List of Figures

List of Tables

List of Listings

1 Introduction

Many articles outline the benefits of *Threat Intelligence* (TI) sharing (McMillan, 2013), (Friedman & Bouchard, 2015), (CERT-UK, 2015), (England, 2016), but only a few provide a rigorous and complete definition. This document aims to provide an in-depth analysis of sharing procedures for TI within the Computer Security Incident Response Team (CSIRT) domain (SA concept level). The aim is also to provide an in-depth analysis of how TI communities are established and maintained today. The analysis identifies a first specification for the improvements proposed by PROTECTIVE to optimise TI sharing. The goal is to establish state of the art sharing mechanisms, as well as sharing mechanisms currently employed by the collaborating National Research and Education Networks (NRENs) and Small to Mid-size Enterprise (SME) CSIRTs.

1.1 Defining Threat Intelligence from Data and Information

TI is a difficult concept to define because it carries different meanings dependent on its purpose and use. In an environment of law enforcement authorities or agencies, TI may relate to information about real-world physical threats (e.g. a planned terrorist attack). The term "*Threat Intelligence*" is often used synonymously with *Cyber Threat Intelligence* (CTI) in the literature, and for our documents, we will continue to do so.

Henry Dalziel (Dalziel, 2014) stated that information must meet three requirements to be classed as TI. It should be **relevant, actionable** and **valuable**. TI is data that has been refined, analysed, or processed in the way that makes it relevant, i.e. it must relate to "you" in a direct or indirect way. Furthermore, that data must be actionable so it must be specific enough to prompt a response, change, action or decision. Lastly, it must contribute to any useful business outcome to be valuable. In a PROTECTIVE context, we consider TI to be relevant information of some intrinsic value and that is actionable in a local cyber-environment (the domain that the NREN or SME is responsible for). TI might, for instance, be a summary of observed behavioural patterns, activities of known botnets, which may take the form of email tickets or it may involve for instance forwarding of an *Intrusion Detection System* (IDS) alert.

Centre for the Protection of National Infrastructure (CPNI) (Chismon & Ruks, 2015), stated that the purpose of intelligence is the process of moving identified topics, in particular risks, in a way coherent with the concept of 'knowns' and 'unknowns'. Different categories (see Figure 1) are used as a coarse classification scheme allowing identification of how much is known about topics under investigation. In principle, with enough intelligence, the unknowns become known eventually, see Figure 1.

Figure 1: MWR and CPNI's description of the purpose of intelligence. Image courtesy of MWR and CPNI (Chismon & Ruks, 2015)

CPNI describe intelligence as: "*… information that can be acted upon to change outcomes. It's worth considering traditional intelligence before exploring threat intelligence, as in many ways the latter is simply traditional intelligence applied to cyber threats*." They also argue that as TI is a young field, there are vendors and advisory papers that describe very different products and activities under the TI banner.

It is important to distinguish between **Threat Data** and **Threat Information**. While both carry content that can be used by analysts, the former relates to lower-level raw logs that have been produced by sensors, whereas the latter will have undergone some additional processing. Threat data such as payloads, Internet Protocol (IP) addresses and Uniform Resource Locator (URL)s are so short-lived that the data may be outdated before reaching their intended reader, but this data can be useful in forensic investigations, if delivered on time, and for correlation and prioritisation purposes. Meanwhile information – i.e. data that has been processed to provide enhanced high-level insight, may help us make well informed.

Intelligence can include human processing (e.g. the TI comes in the form of a new email ticket in an emailing ticketing system), or automated processing (e.g. the TI comes in the form of processing that has happened on the raw logs before being sent out). An example of this is the Meta Alerts that will be described in D3.1 as: "*A meta-alert is similar to an alert, but its contents also include values obtained as results of functions that takes the attributes of the merged alerts as arguments. The purpose of meta-alerts is to aggregate information of related attacks and present a single alert instance that summarizes all the relevant information to a human analyst*".

These examples would suggest that intelligence is more high-level than data (Zeleny, 2005), (Lievesley, 2006) (Rowley & Hartley, 2006) (Rowley, 2007) (Zins, 2007); a recurring example of this perspective can be found in the *Data, Information, Knowledge, Wisdom* (DIKW) Pyramid, see Figure 2. In the context of PROTECTIVE, we do not use the Pyramid to rigorously define how information and data become TI, but as a conceptualisation of how information and data can be differentiated from each another.

The pyramid has spawned a vast amount of literature to discuss its validity. While there are significant disagreements about how to define each level, few works dispute its conceptual and theoretical value (Frické, 2009). In the context of PROTECTIVE, we mainly concern ourselves with the bottom two levels of the pyramid: data and information. Analogously, in PROTECTIVE we can treat information as intelligence (under the assumption that the information satisfied the criteria being relevant, valuable and actionable). Information feeds into analysts' knowledge and wisdom.

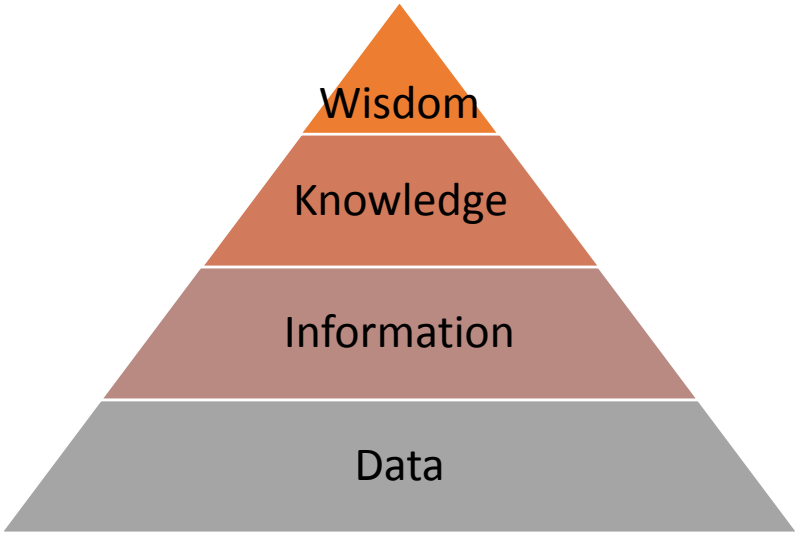

Figure 2: The Data, Information, Knowledge, Wisdom (DIKW) Pyramid (Frické, 2009)

As an example, data would encompass low-level records such as network packets, binary snippets, netflow or IDS alerts. More broadly speaking, data can be regarded as:

- **Facts**: a value that has the fundamental properties of being "*discrete, objective facts or observations, which are unorganised and unprocessed and therefore have no meaning or value because of lack of context and interpretation*" (Rowley, The wisdom hierarchy: representations of the DIKW hierarchy, 2007);
- **Signals**: sensor output and the raw patterns that arise from streams of data. These streams of data can be either subjective or objective in nature;
- **Symbols**: data that takes the form of strings, binary, files, integers etc.

Information can be regarded as: "*organized or structured data, which has been processed in such a way that the information now has relevance for a specific purpose or context, and is therefore meaningful, valuable, useful and relevant*" (Rowley & Hartley, Organizing Knowledge: An Introduction to Managing Access to Information, 2006). TI delivers aggregates or some processed feed that can provide a way for security vendors to use information about threats that go beyond what raw data logs can tell an analyst. In other words, information can be considered as data that has been processed to some degree, whether this is human processed, such as an email being sent from one analyst to another, or aggregates of low-level data that attempt to provide some higher-level insight or interpretation about that data.

Intelligence helps analysts solve high-level problems, e.g. identify common patterns, while data is akin to obtaining logs and fixing the immediately observed issues. There is however, nothing precluding TI to be supplemented by threat data (e.g. logs are accompanied TI in order to provide examples of how the intelligence was generated). Analogously, we can consider the necessary TI capability in CSIRT comparable to how anti-virus software reports virus activity on a desktop computer. Once TI has been reported, the user should be able to, or at least know, what to do with this intelligence to protect their organisation. In an operational sense, it is also important to consider how information feeds into knowledge and wisdom aspects of the pyramid (specifically to identify requirements and specifications of the PROTECTIVE tool, see D2.1). Specifically, knowledge can be considered as:

- **Processed**: *"synthesis of multiple sources of information over time"*
- **Procedural**: the *expertise* in an operational environment – i.e. what to do with information and data in order to make a sound judgement, i.e. *"information combined with understanding and capability".* The purpose of PROTECTIVE is to facilitate well-informed judgements through context awareness and situational awareness.
- **Propositional**: *"the subjective 'perception of the world and one's place in it"* (Boulding, 1955)

Finally, an analyst gains *wisdom* by having made many judgements in the past and learnt from past mistakes. *"Wisdom is the ability to increase effectiveness. Wisdom adds value, which requires the mental function that we call judgment. The ethical and aesthetic values that this implies are inherent to the actor and are unique and personal."* (Rowley, The wisdom hierarchy: representations of the DIKW hierarchy, 2007)

There are several adaptations of the pyramid, with different definitions of Data, Information, Knowledge and Wisdom. Some adaptations also end in Knowledge rather than Wisdom. We use the Wisdom tier to consider how analysts also learn from past experiences. With this understanding of TI and threat data in mind, we can move on to describing how existing literature has defined TI and threat data.

It is worth noting that the literature reviewed in this document does not use the term TI consistently. Some authors use intelligence synonymously with any of the four levels in Figure 2. We use the term TI to refer to relevant and valuable information and/or data that has been structured in such a way it

can be interpreted by a human or a machine to make an action. A breakdown of how PROTECTIVE considers information types can be found in Section 2.2.3.

1.1.1 Existing Definitions of Threat Intelligence

In the first instance, we consider TI to mean information that carry some value to NRENs and SMEs to make actionable decisions, and aids in proactive and reactive handling of cyber-attacks and cybersecurity incidents. However, we also need to consider how others have defined TI outside the context of NRENs and SMEs. We explore existing definitions of TI from a variety of organisations and authors, including (listed in alphabetical order): Bank of England, Bouchard and Friedman, Burger et al., Computer Emergency Response Team – United Kingdom (CERT-UK), Gartner, CPNI, and The SANS Institute.

The Bank of England published a white paper (England, 2016) titled: "*Understanding Cyber Threat Intelligence Operations*" in which they describe intelligence as "*a particular kind of information. Intelligence and information are often used interchangeably as are information and data. To properly understand information (and therefore intelligence) it is necessary to put it in context and a useful model is the data information knowledge pyramid.*" This is the same DIKW Pyramid from the previous section. Their document has highlighted how "*Data and information are often used interchangeably despite being different things. One potential source of confusion is that information can itself be subject to further abstraction and manipulation, in other words, one person's information can be another person's data."*

They follow their description of intelligence up by describing TI as: "*the contextualised output of a strategically-driven process of collection and analysis of information pertaining to the identities, goals, motivations, tools and tactics of malicious entities intending to harm or undermine a targeted organisation's operations, ICT systems or the information flowing through them.*" This definition points out that TI should be able to contextualise data in ways we can use it to protect our organisation strategically.

The authors break down TI into different categories of sources, including:

- **Human Intelligence** (HUMINT): intelligence from people – overtly
- **Covert Human Intelligence Sources** (CHIS): intelligence from people – covertly
- **Open Source Intelligence** (OSINT): intelligence from publicly available sources – overtly
- **Signals Intelligence** (SIGINT): intelligence derived from interception of signals – overtly or covertly
- **Technical Intelligence** (TECHINT): signals captured and interpreted routinely by hardware devices or software applications – overtly or covertly

Unlike other reports, the Bank of England also discuss the lifespan of TI. Their document argues that a TI life-cycle follows a traditional computing-process model of input-process-output (England, 2016). They posit that in this way the intelligence function can match the pace of change in both the intelligence field and the threat environment, see Figure 3. We believe the lifespan of TI will be important to consider because TI is likely to have an expiration date before it loses its intrinsic value to an NREN (i.e. once information is deprecated, knowing the TI will not have a direct effect on the NRENs ability to prevent or respond to the threat). For instance, an IP address that belonged to a botnet a month ago, may have been cleared since it was last observed.

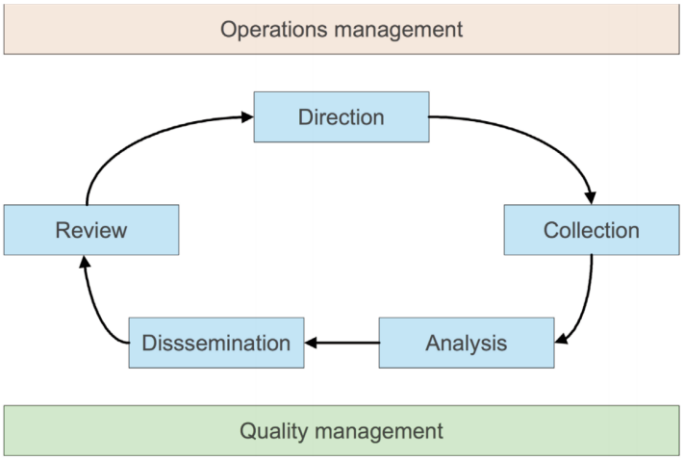

Figure 3: Lifespan model of TI. Image Courtesy of Bank of England (England, 2016)

Friedman & Bouchard (Friedman & Bouchard, 2015) define TI as: "… *knowledge about adversaries and their motivations, intentions, and methods that is collected, analysed, and disseminated in ways that help security and business staff at all levels protect the critical assets of the enterprise*." They also provide several key characteristics of their definition, including how TI can be:

- **Adversary-based**, akin to military and police activities attempting to stop an enemy of the nation, but also in the competitiveness sense found in sports.
- **Risk-focused**, i.e. based on an assessment in terms of assets that need protecting.
- **Process-oriented**, protect the organisation mission flows of activities.

A challenge with simply adopting this definition is that it makes no assumption about the type of knowledge is being referred to. Procedural knowledge is for instance difficult to provide evidence for, while formal, codified, or explicit (i.e. declarative) knowledge however may be more straightforward to automate. In other words, knowledge may be too high-level to document and communicate.

Burger et al. (Burger, Goodman, Kampanakis, & Zhu, 2014) do not provide a definition of TI, but present a layered taxonomy model for cyber TI sharing technologies. The model follows potential technology options and instantiations rather than a strict hierarchical model. The layers, shown in Figure 4, are akin to the ISO OSI protocol model, carrying ideas such as transport, session, indicators, intelligence, and 5W's (who, what, where, when, and why).

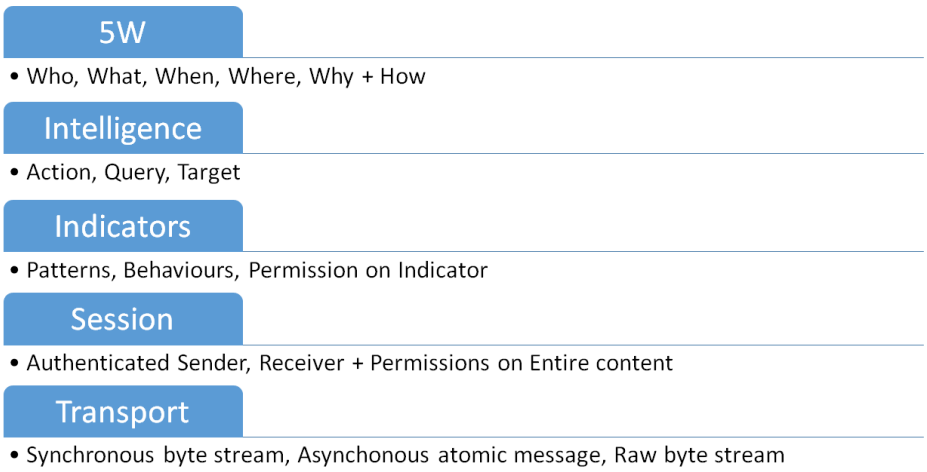

Figure 4: The layered taxonomy model by Burger et al. (Burger, Goodman, Kampanakis, & Zhu, 2014). Image drawn after Burger et al.

CERT-UK, in their article "*An introduction to threat intelligence*" (CERT-UK, 2015), use Gartner (see below) definition to describe TI as: *"… 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".* This definition proposes an element of traceability (evidence) has to been in place for any decision-making capability from the information provided. The definition also relies on the TI being processed information to make cognitive assessments about the insight provided. The type of knowledge is not specified.

CERT-UK continue by describing how TI is generated and can be deployed at strategic, operational and tactical level, as shown in Figure 5. They follow on to describe the role of TI to be means to aid *Security Operation Centres* (SOCs) in triage and responding to security-related incidents.
The document continues by advocating Structured Threat Information eXpression (STIX) and Trusted Automated eXchange of Indicator Information (TAXII), by stating that the purpose of STIX/TAXII is to "*automatically create indicators of compromise (IOCs) to inform rules in their intrusion detection systems (IDS), intrusion protection systems (IPS) and firewall devices. From a management perspective however, the Threat Intelligence Platforms (TIP)s must simultaneously provide near real-time tactical threat intelligence, inform operational concerns for handling such threats, and support strategic prioritisation of cyber-security issues across the organisation.*" The report follows on to describe that TI services should also provide information tailored to the client's network; prioritising vulnerabilities, predicts threats and enable fast response.

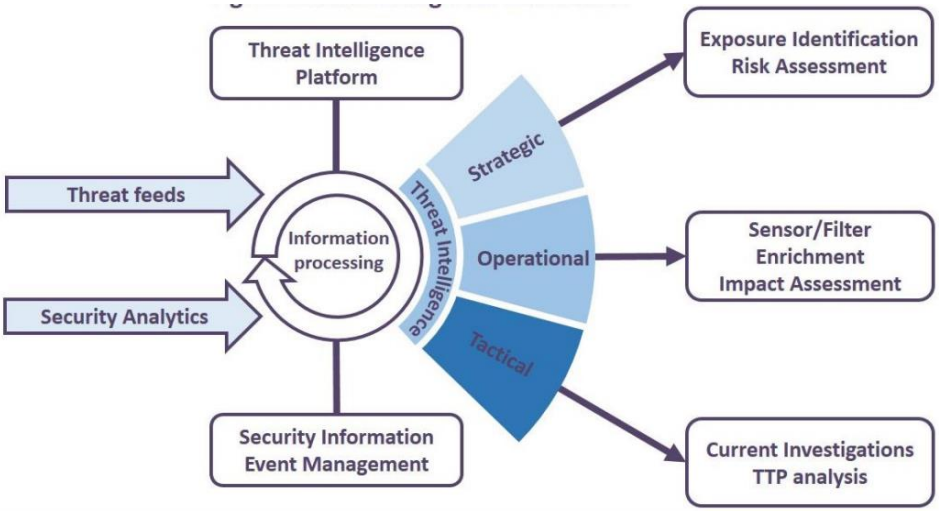

Figure 5: Processes involved in generating TI. Image courtesy CERT-UK (CERT-UK, 2015)

European Union Agency for Network and Information Security (ENISA) has a number of documents outlining the importance of TI as well as how it can be used (ENISA, 2013) (ENISA, 2014) (ENISAa, 2014) (ENISA, 2016) (ENISAa, 2016). While no dictionary-like definition has been provided, ENISA (ENISAa, 2014) proposes a distinction of the various *actionable information* types that can be observed related to threats. Their approach is based on four-layers of information that includes: *low-level information*, *detection indicators*, *advisories* and *strategic reports* (see Section 2.2.3). ENISA has also provided an in-depth threat taxonomy (ENISA, 2016) consisting of:

- **High-level threats**: top-level threat category, used mainly to discriminate families of threats.
- **Threats**: this field indicates the various threats within a family.
- **Threats details**: detail specific threats. Typically based on a specific attack type/method or targeting specific asset.

Gartner (McMillan, 2013) defined three levels of cyber TI that CERT-UK also use:

- **Tactical: technical intelligence such as using threat indicators to proactively hunt for and defend against adversaries;**
- **Operational**: intelligence focused on the motivations intent and capabilities (including TTPs) of adversaries
- **Strategic**: intelligence about the risks and implications associated with threats used to inform business decisions and direct cyber security investment.

Other vendors that have adopted Gartner's definition include Tripwire (Tripwire, 2014), FireEye (FireEye, 2017) and Webroot (Webroot, 2014). Webroot points out that TI is not: 1) obvious, trivial or self-evident information about a threat that individuals can discern for themselves; 2) (purely) information about vulnerabilities; and 3) support for incident response.

CPNI's report on TI (Chismon & Ruks, 2015) describe TI as a new field in cybersecurity that applies the notion of traditional intelligence with cyber threats, and defines TI as belonging to one of four main subtypes: Short term use, Long-term use, High-level and Low-level TI, see Figure 6. This is similar but not identical to the literature (especially Gartner) due to the presence of *operational*, *strategic* and *tactical* levels – here CPNI have introduced the technical level, which is intended to be the short-term, low-level indicators of specific malware or other nefarious actors. In PROTECTIVE, our definition of TI (see Section 1.1.2) and threat data encompasses the information in each of these quadrants.

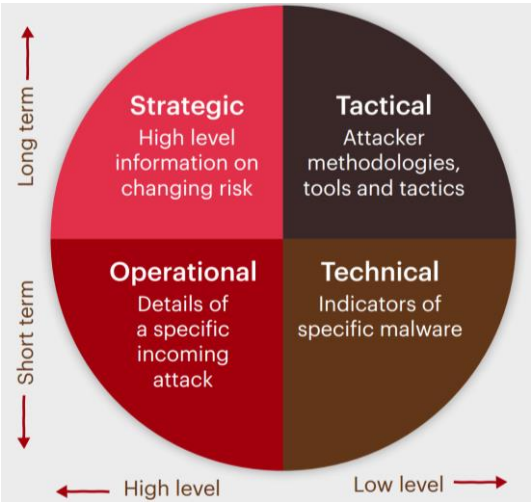

Figure 6: CPNI's categorisation of cyber intelligence. Image courtesy CPNI (Chismon & Ruks, 2015)

As with traditional intelligence, the report argues that: *"a core definition is that threat intelligence is information that can aid decisions, with the aim of preventing an attack or decreasing the time taken to discover an attack. Intelligence can also be information that, instead of aiding specific decisions, helps to illuminate the risk landscape"*. The latter part of this definition falls in line with Bouchard and Friedman's description in that TI also can provide added insight into the risk landscape, and not only focus on threats and threat actors (which most definitions appear to aim for).

The SANS Institute description does not provide a working definition of TI, but instead adopt Gartner's definition (Bromiley, 2016) and describe what TI can help with considering aspects of cyberattacks; for example, how TI can share data and information such as being able to help victims identify information about delivery mechanisms, infrastructure affected, as well as motivations and the actors involved themselves, see Figure 7. Furthermore, they explain that: "*Part of defining TI is deciding what it is not. TI is not simply a list of atomic indicators that an attacker used at one point in time, without additional context into how the attack worked.*"

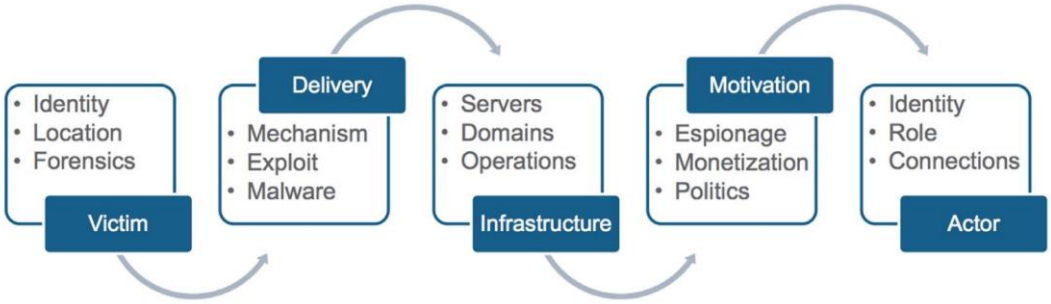

Figure 7: TI provides information about any of the listed components of an attack. Image courtesy of SANS (Bromiley, 2016).

Finally, National Institute of Standards and Technology (NIST) (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) simply define TI as: *"Threat information that has been aggregated, transformed, analyzed, interpreted, or enriched to provide the necessary context for decision making processes."* This definition points to TI being information about threats that is necessary for context in decision making. Comparatively speaking, this is a somewhat rudimentary definition in that it does not explicitly mention data, or that it is evidence-based. This raises several concerns in that without evidence, TI could potentially be (for instance) an email being sent about a person's instinctual reaction, and this would still count as TI because it is interpreted information that provides context to decision making processes.

1.1.2 PROTECTIVE Threat Intelligence Definition

The project is aligned with ENISA's definition of threat intelligence (ENISAa, 2014), in order to contextualise it with a dictionary-like definition, we informally describe TI similarly to Gartner, as: "*evidence-based information and data including insight about context, mechanisms, indicators, implications and actionable advice, about a past, existing or emerging menace or hazard to assets or processes that can be used to inform decisions regarding the organisation's response to that menace or hazard."*

TI can be comprised of Data and Information. Data are elementary facts and observables, while Information is data in context, or a higher-level abstraction or viewpoint based on one or more data items (akin to that described by Bank of England (England, 2016)). The key difference is that we make a distinction between low and high-level insight and classify them as data and information, we also assume a working example, based on the *Intrusion Detection Extensible Alert* (IDEA) schema[1], see section 2 for an example.

[1] https://idea.cesnet.cz/

2 TI Sharing – Concepts, Building Blocks, and Case Studies

TI sharing is becoming increasingly important to organisations today (McMillan, 2013), (England, 2016) (CERT-UK, 2015), (England, 2016). It can enable them to improve automation (where appropriate), enhance insight about patterns of threats, provide context for their own alerts, aid patching, help improve proactive and reactive strategies (incl. policy creation), learn lessons from the misfortune of others, reduce false positives and stay up to date about the current threat landscape. In the following sections, we outline the various concepts, building blocks and case studies that exist and will help address these challenges, before moving onto the requirements and specifications. We also reflect on a number of case studies from the literature to aid our requirements.

2.1 TI Sharing Concepts and Factors

The main purpose of sharing TI is to aid other organisations in defending against attacks. A number of challenges exist to support benefits of TI sharing, including legal, technical and efficiency challenges. We outline key literature on these sharing factors below and propose how to address these in our requirements and specification section, see Section 3. The purpose of this section is to aid understanding of best practices, as seen from the state of the art.

Sillaber et al. (Sillaber, Sauerwein, Mussmann, & Breu, 2016) investigated how a number of organisations have an increased willingness to participate in TI sharing platforms. They conducted focus group discussions with ten stakeholders from SOCs. The study identified several factors that affect shared TI data quality at multiple levels, and presents limitations and complexities associated with integrating and consolidating shared TI from different sources while ensuring the data's usefulness for an inhomogeneous group of participants. These factors include: 1) Integration of TI sources amplifies pre-existing data quality problems; 2) Combining short-lived shared TI from different industries makes the important intelligence hard to find; 3) Existing TI sharing tools often limit data accessibility; 4) Manually generated quality errors are difficult to find and often occur due to a lack of common data entry rules; and 5) Automated integration of external sources can improve data quality.

The study found that there are no fundamentally new data quality issues that are unique to TI. However, TI is an emerging new domain with several tools being rushed into market, with many integration issues having not been addressed. The authors propose several recommendations, including:

- ensuring that the TI sharing standard and meta-model fit the stakeholders' needs.
- minimising complexity by unifying intelligence according to vulnerabilities.
- correcting data at the main source and link intelligence at the borders.
- trust in data is of utmost importance: inform users about data quality – including esp. provenance, verifiability (of data).
- automating data quality error detection.
- crowdsourcing data quality management – including reputation of the data.
- focusing on current TI as information gets outdated over time.

Ahrend et al. (Ahrend, Jirotka, & Jones, 2016) did an empirical study to identify intangible collaborative practices that cybersecurity analysts rely to "do" TI. The authors attempt to unpack the informal forms of collaboration and coordination at work that build tacit knowledge about threat actors and defenders across time, people and tools that help turn threat information into TI. Semi-structured interviews and diary studies were conducted at three TI service providers with five participants. The authors highlight limitations in the human psychology side of TI by introducing the concept of "*Threat and Defence Knowledge*" (TDK). This is tacit knowledge that analysts within an organisation form over time and utilise through informal ways (including correlating it with other data). The authors believe

that the lack of access to this knowledge can reduce analysts' effectiveness in acquiring TI and therefore reduce the ability to perform.

Perceived and potential shortcomings of the existing processes and tools include:

- Awareness relies on casual day-to-day interactions and informal monitoring and overhearing.
- Awareness is limited to teams and departments engaged due to the dependency of passive overhearing on physical proximity.
- A decrease in TDK utilisation is potential loss of tacit knowledge, stemming from an unavailability of TDK originators (e.g. change of team, department- or organisation-affiliation) or memory loss.
- TDK artefacts on the other hand can be documented, but achieving their availability and obtaining the necessary access rights can be difficult.
- TDK artefact's context is at times documented but again often restricted to the originator's personal machine. TDK sometime therefore reply on manual queries of an analyst and the TDK originator's ability respond accurately and holistically.

Ahrend et al. argue that each analyst in effect act like a database in which incidents and documents are linked together, which is something we believe PROTECTIVE will have to leverage. Analysts' reliance on past experiences and staff turn-over can limit the potential of in-house ability to produce TI in the most effective way possible, especially with a significant amount of knowledge being tacit and undocumented. The authors argue that traceability and contextualisation of past investigations is important when discussing TDK. These are issues that the sharing model will have to address.

Garrido-Pelaz et al. (Garrido-Pelaz, González-Manzano, & Pastrana, 2016) present a model to analyse the benefits and drawbacks of information sharing among organisations that present a certain level of dependency. Their model applies functional dependency network analysis to emulate attacks propagation and game theory for information sharing management. They present a simulation framework implementing the model that allows for testing different sharing strategies under several network and attack settings. Experiments using simulated environments show how the proposed model provides insights on which conditions and scenarios are beneficial for information sharing.

Vasek et al. (Vasek, Weeden, & Moore, 2016) present an observational study to measure the short-term and long-term impact of sharing abuse data with web-hosting providers. Their dataset of URLs blacklisted for distributing malware comprises over 28 000 URLs shared with 41 organisations. The authors show that sharing has an immediate effect of cleaning the reported URLs and reducing the likelihood that they will be compromised again. The authors claim that they found limited evidence that one-time sharing of malware data improves the malware clean-up response over the long term. The study does not go more in-depth into why this might be the case (presumably due to the data not being available).

2.1.1.1 Reflection on Sharing Concepts and Factors

From the literature, we see several emerging themes among factors that need to be addressed within the PROTECTIVE in terms of TI sharing. These include questions linked to factors relating to policy, jurisdiction, trust, psychology, human-computer interaction, operations, managerial tasks, context, resources, lack of universal agreement on what TI is as well as factors relating to technical limitations.

With policy and legal factors, it becomes necessary to ask about organisation insight into national and international policies and jurisdiction. For instance, has there been a lack of knowing about a policy or legal jurisdiction? Have issues emerged in looking up whether sharing or responding to TI is a problem? Have NREN ever been not able to share because of current legal/policy rules? Is this likely to be a problem in the future?

Trust is important in TI sharing. On one hand, it can be related to the transmission of TI, e.g. assurance that TI is sent from and received by appropriate sender and recipient, respectively. However, it can also be relate to confidence in the TI quality sent. Levels of confidence might relate to accuracy, precision, completeness, timeliness, freshness, and relevance of TI, or how much confidence an organisation should have in a sender (based on their prior history, i.e. "*reputation*").

People are interpreters of TI at the higher level. Some human-related elements make TI sharing a challenge, and it is necessary to identify these. For instance, are there any work-culture, national, cultural factors that have an impact on when some TI is shared or not? A large part of this relates as well to TDK that is used across an organisation. A part of this also includes the elements of Human-Computer Interaction (HCI), i.e. how people and machines work together. It is, therefore, necessary to identify elements in HCI that limit, or even further enable, analysts' work capabilities.

Operational and managerial factors relate to day-to-day activities, some of which may facilitate or obstruct appropriate TI sharing. For instance, an important report due while an incident is happening means that the incident may have to be dealt in a later moment of time. As part of the operational aspect, there may be contextual factors that can influence TI sharing. For instance, analysts already familiar with them can deal the repetitive scenarios straightforwardly. New scenarios, that are unfamiliar for the analyst, may result in slower response times, human error or misinterpretations. Moreover, scenarios that are brand new or driven by particular mission contexts can also influence TI sharing. Important to consider are resources relate to operational factors, specifically challenges related to manpower, budgetary issues or limitations in technology constraints. Finally, a lack of standardisation in the TI sharing space may also contribute to challenges in this space.

2.2 Building blocks for Threat Intelligence Sharing

Threat intelligence sharing systems are composed of a number of functional elements or "building blocks" which include:

- **The models and modes of TI exchange**, i.e. with whom the information is shared and how? What is the impetus behind information sharing? Is it shared voluntarily or a regulated requirement?
- **Rules of engagement and protocols**, i.e. what are the agreements, rules and procedures in place to share TI responsibly?
- **Types of information exchanged**, i.e. what information is being shared, and what is the purpose of sharing it?
- **Mechanisms of exchange**, i.e. How is the information actually shared?

We examine each of these in more detail in the following sections.

2.2.1 Models/Modes of TI Exchange

2.2.1.1 Communities

Organisations are continuously under threat from cyber-attack and it is difficult for them to have a total overview of all threats. Sharing threat intelligence with similar organisations through the medium of TI-sharing communities can help improve an organisation cyber defences.

(Fransen, Smulders, & Kerkdijk, 2015) define a TI community as a "network of organisations that start exchanging threat intelligence amongst each other", while (Willis, 2012) identifies an information sharing community as a "group of trusted stakeholders who work together to address shared threats or vulnerabilities". Willis also suggests that a community may take the form of a public-private partnership or industry-to-industry partnership.

TI-sharing communities can form for various reasons (Willis, 2012), (Serrano, Dandurand, & Brown, 2014), (Wagner, Dulaunoy, Wagener, & Iklody, 2016), (TeleManagement Forum, 2013), (Harkins, 2016) which include:

- **Enhanced depth and breadth of insight**: allowing an analyst at one organisation to share information with an analyst at another in order to enhance threat understanding, knowledge maturation and greater defensive agility.
- **Confidentiality assurance**: ensure organisations that sensitive information is being handled properly with proper control over communication paths.
- **Common interests**: establish ability for analysts to reach other analysts to find a solution to the same threats that they are facing.
- **Bigger picture awareness**: monitor changes in the threat landscape, which helps in identifying the potential cyber threats that can cause disruptive effect on society. This bigger picture view may help create a new situational awareness that adds value for the whole community.

The great diversity of industries, companies, public entities and other potential sharing partners coupled with the range of TI information-types raises the question of what to share, how to share and with whom to share. NIST (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) describe communities as informal or formal. An informal community is an open, self-organising group that operates under basic rules of conduct without a formal agreement. In contrast, formal communities have specific membership rules whereby group members are vetted and where exchange of information is often governed by service level agreements, non-disclosure agreements and other agreements that describe member's responsibilities.

It quickly becomes apparent that there is no "*one size fits all solution*" for TI sharing and that a variety of sharing community types will in practise be needed. Factors shaping communities include security goals e.g. per industry sector, the level of confidentiality of the information, organisational trust models, geographical locations and so on. A tiered engagement model arises naturally in response to these needs (Willis, 2012), (Harkins, 2016), (Wagner, Dulaunoy, Wagener, & Iklody, 2016), (Zhao & White, 2012). In this model the lower tiers are more expansive in terms of membership and information distribution and the upper tiers are more specific with a narrow security focus and, generally, more selective membership.

Based on the model of (Willis, 2012) we propose the following tiered information sharing model – see Figure 8 - and which is described further in Table 1.

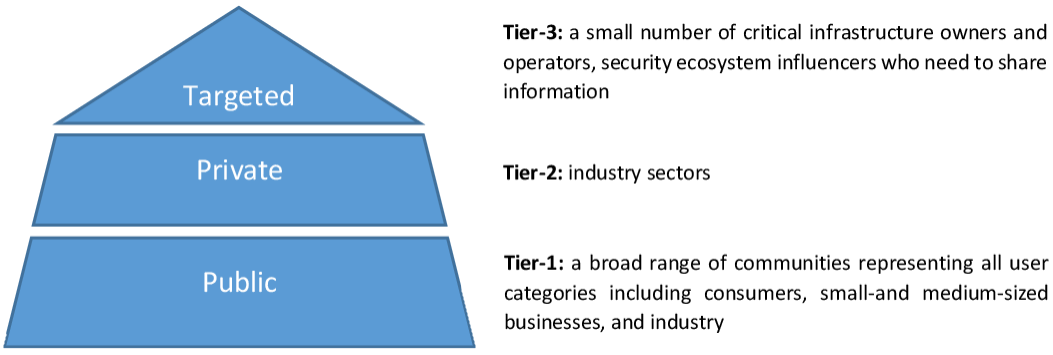

Figure 8: Proposed tiered information sharing community model (adopted from Willis (Willis, 2012))

The model consists of three tiers as follows:

- **Public tier (Tier-1)** – this tier consists of communities with greater trust and low level of sensitivity. Community membership is informal and voluntary in Tier-1. Information is broadcasted to all (users, consumers, small- and medium-sized businesses etc.). The Tier-1

community structure is equivalent to Willis's public tier and Serrano's exchanges containing public information (Serrano, Dandurand, & Brown, 2014).

- **Private tier (Tier-2)** – this consists of communities that desire to restrict information sharing for security, commercial or other reasons. It includes both private and public sector organisations including regional, national, international, sector-specific or common interest driven. Membership may be either informal or formal. Examples from the literature include Willis's confidential tier (Willis, 2012), Serrano's private knowledge exchange (Serrano, Dandurand, & Brown, 2014), Wagner's internal, community or national classification schemes in Malware Information Sharing Platform (Wagner, Dulaunoy, Wagener, & Iklody, 2016), Zhao & White's sector or non-sector organisations (Zhao & White, 2012) .
- **Targeted tier (Tier-3)** – this consists of focused communities' specific needs on the type of information shared or with particular security requirements such as critical infrastructure industry sectors. Tier-3 community structure is equivalent to Willis's targeted tier (Willis, 2012).

The principal characteristics of these tiers is described further in the table below:

	Community structure	Goal Example	Framework
TIER-1: PUBLIC	Consists of broad range of communities that represent all user categories, including consumers, small-and medium-sized businesses, and industry in general	Forum for Incident Response and Security Teams (FIRST).[2]	Trusted frameworks are not necessary. Communities typically distribute information broadly through mechanisms such as e-mail distribution lists or public web sites
TIER-2: PRIVATE	Consists of communities that represent industry sectors or other common interest driven groups	NREN's might share the command and control internet addresses that botnet use; members can quickly identify and block botnet sites	Communities typically use trust frameworks such as non-disclosure agreements or memoranda of understanding
TIER-3: TARGETED	Consists of small number of communities, and member organisations who need to share critical information e.g. for protecting critical infrastructure	Information Sharing and Analysis Centers (ISACs)	Strong information-sharing framework, such as national security clearances and non-disclosure agreements, are required. Trusted sharing mechanisms, such as encrypted web portals with multifactor authentication, may also be required

Table 1: Characteristics of proposed tiered information sharing community model (Adopted from Willis (Willis, 2012))

NREN sharing models today are usually of type Tier 2 where an NREN community is typically composed of the NREN itself and its primary constituents. These communities are informal and the types of information shared includes constituency specific threat information as well as general advisories. A related example in this area is Malware Information Sharing Platform (MISP) (Wagner, Dulaunoy,

[2] FIRST - Information Exchange Policy framework Version 1.0
https://www.first.org/iep/FIRST_IEP_framework_1_0.pdf

Wagener, & Iklody, 2016). MISP relies on the voluntary action of its community to share information and indicators. Furthermore, the level of reach of the content is left to the sharer, who can select various sharing scenarios, as described below:

- **Organisation only**: Only members of an organisation are allowed to see an event.
- **Community only**: Users of the MISP community can see the event, including organisations that run MISP servers that synchronize with that server.
- **Connected communities**: Users of the MISP community including organisations on this MISP server, as well as MISP servers synchronizing that server. This also includes hosting organisations of servers that connect to these servers.
- **All:** The shared content is shared within the whole MISP communities.
- **Sharing group**: A distribution list approach that can include a set of organisations and remote MISP instances. This setting allows for granular distribution as well as the option to entrust partners with an extending role within the sharing group.

The above categorisation introduces the question of sharing between communities as well as within communities. An example of the inter-community sharing is the North Atlantic Treaty Organisation (NATO) MISP connected community shown in Figure 9 below. This shows that two organisations (NATO and BEL MOD) have MISP instances and can synchronize information. Other organisations (CERT-BW, TUBITAK, DefCERT NL, FCCU, and BELNET CERT) can connected to the organisation having MISP instances for contribution and information collection.

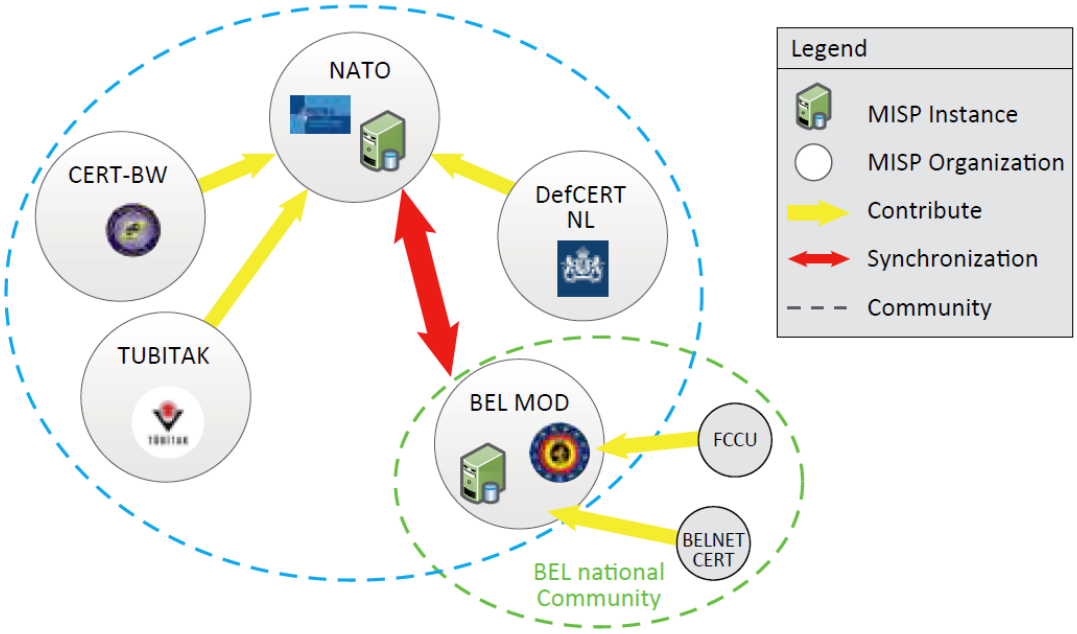

Figure 9: NATO MISP Community. Image courtesy NCIA (NCIAgency, 2017).

Another model of inter-community sharing is that of Serrano et al. (Serrano, Dandurand, & Brown, 2014) who proposed the idea of a *Knowledge Exchange* (KE). A Knowledge Exchange (KE) is "*a service containing a list of data publishing organisations and their associated data and/or service offering*" (Serrano, Dandurand, & Brown, 2014). Public and Community KEs allows inter exchange communication. A public KE allows querying known organisation for new exchanges and asking known exchange for new organisations whereas a community KE will exchange information with other exchanges based authentication information. Serrano also defines a Private KE but this shares information only within a community and not between communities.

To establish trust when sharing information in a community, there is a need to have secure access. According to Multinational Alliance for Collaborative Cyber Situational Awareness (MACCSA, 2013), as

businesses become more collaborative, there are increasing requirements for accountability and information protection. The TM Forum (TeleManagement Forum, 2013) elaborates on the need for access control, which comprises the following:

- **Identity:** A way of vetting new members or organisations wishing to join the community is required. A mechanism for viewing the profile of members with whom new or existing members may want to connect should be provided.
- **Authentication**: both users and machine-to-machine communication is verified either through username and password, or mutual certificate based authentication. There is also a need for a trusted certificate authority.
- **Authorisation**: Users are authorised access to different levels of information being shared in the community based on a common understanding of community roles.

2.2.1.2 Organisational Trust Models

TI information-sharing can range from ad-hoc exchanges to exchanges established through mid-term or long-term formal agreements. The different approaches reflect characteristics such as the level of trust between the sharing parties, the relationships between the stakeholders, and the legal authority of various actors. For organisations to obtain levels of trust prior to sharing information, NIST 800-39 provides five trust models (validated, direct historical, mediated, mandated and hybrid) but also notes that "*no single trust model is inherently better than any other model. Rather, each model provides organisations with certain advantages and disadvantages based on their circumstances*". These are detailed below.

Validated Trust: A first organisation uses a body of evidence on a second organisation to establish a level of trust with the second organisation. The greater the degree and quality of evidence provided between two organisations, the greater the level of trust. However, the amount of evidence provided might not be enough to fulfil the trust requirements of the first organisation. In addition, obtaining the required evidence might not be possible.

Direct Historical Trust: In this model, the history of an organisation in the past helps establish a level of trust. Trust may be based on an organisations risk and security-related activities and decisions, the organisations working relationship with each other on any other activity. In other words, trust needs to be built up over time.

Mediated Trust: A mutually trusted third-party organisation provides assurances that lead to a trust level between two organisations. The concept of transitivity of trust whereby A trusts B, B trust C, therefore A can trust C. For example, if two pairs (A, B) and (A, C) of NREN organisations have already established relationships and trust, NREN A can introduce B and C to each other.

Mandated Trust: A level of trust is established through a specific mandate issued by a third party in a position of authority. This requires that some organisational entity is decreed to be the authoritative source of the shared information.

Hybrid Trust: This is a combination, of any or all, of the above for an organisation. Also, larger organisations may use different trust models in different departments or sub-organisations. One example of an NREN's hybrid trust model involves collecting specific information (validated) prior to attaching sharing hardware; being the certificate authority for issuing certificates (validated/mandated) and being the provider of the TI to all connected organisations (mandated).

Similarly, Microsoft (Goodwin, et al., 2015) presents four commonly used ways of information exchange as:

- formalised exchanges;

- security clearance based exchanges;
- trust-based exchanges;
- ad-hoc exchanges.

These correspond, in part, to the NIST classification with ad-hoc being the notable exception. (Zhao & White, 2014) also provide a formal model for collaborative information sharing for cybersecurity community. The MACCS Information Sharing Framework (MACCSA, 2013) which does not distinguish between formal and informal communities, identifies the need for information sharing agreements between members. According to Microsoft (Goodwin, et al., 2015) there are two information-sharing models:

- **Voluntary model:** In voluntary model of sharing, actors identify a need or reason to share data and usually share valuable and actionable data. It is often decided with whom to share information based on the business or organisation type and the objectives of the parties.
- **Mandatory disclosure model:** Governments, based on their geographical borders, increasingly require the disclosure of threat and incident information to different stakeholders including other government authorities, investors, and customers. To gain access to a community may involve adherence to basic rules of conduct or entering into a formal fixed membership. The idea of informal and formal community membership is presented by NIST (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016).

2.2.2 Rules of Engagement

When a TI community model is established and its type of membership identified, rules that govern the exchange of TI may be required. According to NIST (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016), organisations should establish such rules prior to sharing threat information. Several elements for both the aspects of community membership and handling of the TI are listed by (Serrano, Dandurand, & Brown, 2014) who propose that all data exchange should be governed by means of an *Information Exchange Policy* (IEP)[3]. An example of Machine-readable IEP events are listed in Section 3. Dandurand (Dandurand & Serrano, 2013) suggests the need to define a number of IEPs according to exchange and quality requirements. Such policies establish the protocols or rules of engagement.

Policies according to NIST (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) should be re-evaluated on a regular basis. There should be an agreed procedure for modifying the policy. Triggers that cause a policy change include changes to regulatory or legal requirements, organisation policy and information ownership. Regulatory or legal requirements are rules to ensure compliance with, for example, the *General Data Protection Regulations* (GDPR) coming into force May 2018, and current data protection legislation must also be followed when sharing information.

Considering regulatory requirement and by combining NIST's (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) list of sharing rules and the list of elements from (Serrano, Dandurand, & Brown, 2014)'s IEP, we propose an organisation policy that can be grouped into policy for community membership, sharing information and information ownership:

- **Community membership policy:** For a more formal community membership, organisations might:
  - identify approved recipients of information;
  - agree a procedure for admitting new participants;
  - provide instructions for handling received data for leaving participants.
- **Sharing information policy:** For shared TI, organisations should agree on:

[3] FIRST - Information Exchange Policy framework Version 1.0
https://www.first.org/iep/FIRST_IEP_framework_1_0.pdf

  - the scope of shared information;
  - minimum quality of sharing TI;
  - conditions when sharing is permitted;
  - anonymization/sanitisation of exchanged TI;
  - whether source attribution is permitted;
  - handling requirements and uses that can be made of the exchanged TI;
  - whether exchanged TI can be modified by recipients and subsequently forwarded and obligations for protection of the TI.
- **Information Ownership policy** includes**:**
  - instructions for handling received data of leaving participants;
  - intellectual property rights;
  - agreed retention procedures of the exchanged data.

The *Traffic Light Protocol*[4] (TLP) (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) developed by US-CERT uses a colour coded scheme to indicate expected information sharing recipients. There are no checks or controls to assure the TLP only reaches its intended audience as it is up to the sender to send it in the right direct. CERTs however have used this approach to flag email content in such a way that it is very difficult to unknowingly read material when the intended audience is clearly highlighted. Four different colours are used to represent four levels of sharing:

1. Red - for named recipients only
2. Amber – for particular groups of people
3. Green – for a particular community
4. White – for public sharing

As legislation and regulation evolve to **incentivise data protection**, so too does the case for demonstrating compliance of data management both in spirit and in practise. Such compliance will likely become a necessary component of security and data governance and operations, in order that NRENs and SMEs might manage the risks to which they and end-users share TI. There has been limited research into how to monitoring and managing of TI and onward sharing of data that may or may not include personal data.

Standards bodies, legislation and regulation have outlined data sharing-related concerns that must be considered. Examples include the OECD (O'Leary, Bonorris, Klosgen, Lee, & Ziarko, 1995), European Union (EU) (Movius & Krup, 2009), and United Kingdom (UK) Data Protection Act[5]. Updates to the EU regulations regarding the treatment of data are making this requirement more explicit, particularly with the enforcement of the *General Data Protection Regulation* (GDPR) in May 2018. These concerns are justified by the growing body of evidence that data sharing can put persons at risk to malicious threat (Mauro & Stella, 2016).

The current common practice (in terms of personal data management) is to seek blanket consents with limited ability for change from end-users. This is often more accepted by personal users in the context of service providers for personal use of e.g. social media, simply because the services themselves are so entangled to the personal data in question that is almost a requirement in order for the service to function (although how the data is processed behind closed doors for advertisement and other purposes is another question entirely). Personal end-users often have to accept the use of all personal data, or not be allowed to use the service at all.

[4] https://www.us-cert.gov/tlp
[5] http://www.legislation.gov.uk/ukpga/1998/29/contents

Mauro and Stella (Mauro & Stella, 2016) outline the legal instruments and restrictions for sharing data while complying with the EU data protection law. They summarise that sharing of personal data other than those for which data were originally connected is not forbidden by the GDPR, but on the contrary, it can be carried on the basis of various legal grounds. In particular, they discuss a compatibility test, i.e. ensuring that the handling of data is within the bounds of expected (compatible) behaviour or compliance. This includes the possibilities of how data can be linked, placed in a context, impact and safeguarding procedures, handling personal, non-personal data through informed consent and revocation, but also how formal agreements (e.g. Non-Disclosure Agreement (NDA)) should be in place where appropriate. It is worth stating that the authors also specify there is no definition of "*sharing of data*" under the EU Data Protection Law, including the Regulation (EU) 2016/679 of the European Parliament and of the Council 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. Some EU data protection authorities, which have provided guidance on data sharing agreements, define "*data sharing*" as the "*disclosure of the data by transmission, dissemination or otherwise making it available*" in many different contexts, i.e. within the public or private sectors, or among the public and/or private organisations.

2.2.3 Information Types

A substantial amount of work has been conducted during the last years, towards analysing, discussing and creating a taxonomy of the different information types in the context of TI sharing (ENISAa, 2014) (Goodwin, et al., 2015). Regardless of the chosen taxonomy, the common understanding of the state of the art is the fact that a multitude of heterogeneous types of information exist, which need to be mapped into an expressive class. In Section 1 we discussed threat intelligence in the context of the (conceptual) DIKW Pyramid. Here we break down information to the constituent components in tangible form.

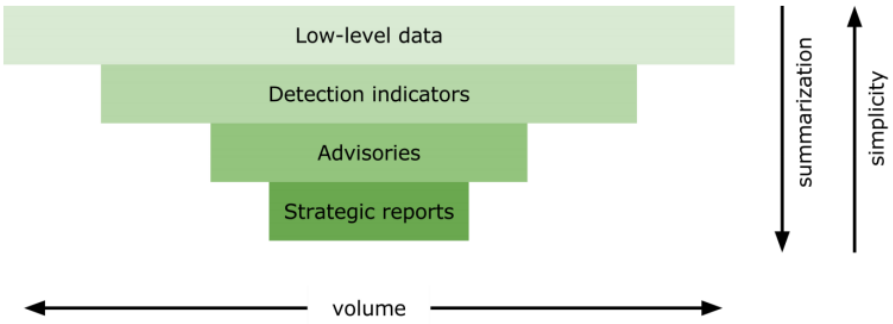

Figure 10: Types of information. Image courtesy of ENISA (ENISAa, 2014)

In this context, ENISA (ENISAa, 2014) proposes a distinction of the various information types that can be observed. It is based on a four-layer approach that includes: *low-level information*, *detection indicators*, *advisories* and *strategic reports*. In the following we discuss briefly each of the four layers. We consider lower level data to be mapped directly to Threat Data described in Section 1, whereas the remaining three levels make up Threat Information. Combined they can make up TI.

2.2.3.1 Low-level data

Low-level data refer to the data collected and generated by various monitoring systems. Examples of low-level data may include: IDS alerts, firewall logs (including flow data and/or full packet captures), application-level logs (e.g. server log files), and operating system-level logs.

It is important to note that, in their majority, low-level data are not useful without additional context. For instance, records of flow data at first glance may not reveal any meaning, while after an analysis

of an administrator they might lead to the detection of an attack. In addition, nowadays, the amount of generated data is so large that makes it infeasible for an administrator to (manually) cope with.

2.2.3.2 Detection Indicators

Detection indicator is defined as a pattern that can be mapped and matched against low-level data to detect malicious activity (ENISAa, 2014). While detection indicators might appear to be similar to low-level data, there is a major difference which is the *context*. An IP address, alone, does not provide enough intelligence; however, an IP address marked as malicious for a specific type of attack does. Examples of detection indicators may include:

- malicious IP addresses (e.g. of infected machines),
- hashes of a malicious entity (e.g. malware),
- honeypot alerts,
- IDS alerts[6],
- URLs of malicious websites (e.g. hosting malicious files).

2.2.3.3 Advisories

The third layer of information types is advisories. This class includes all types of information that are in a higher level than low-level data and detection indicators, that are still actionable; i.e., it can assist in the threat detection process. Examples of advisories may include:

- **vulnerability advisories** (e.g. reports that include additional context such as attack samples, mitigation techniques, etc.);
- **high-level alert data** (e.g. an alert that requires manual interpretation); and
- **adversaries' trends and techniques** (e.g. details regarding the behaviour of a malicious entity).

2.2.3.4 Strategic reports

Strategic reports are high-level summaries usually written in the context of assisting policy-makers during various decision making process (e.g., identifying future attack trends). The level of abstraction of such documents is so high that is considered out of the scope of this document and also not relevant to PROTECTIVE.

Beyond the aforementioned classification, Microsoft (Goodwin, et al., 2015) proposed a taxonomy that follows the pyramid scheme of Figure 11. In more details, and starting from bottom to top, *best practises* refer to software and service security controls, respond practises, etc. *Vulnerabilities* refer to weaknesses in software, hardware (or processes) that may be exploited. Moreover, *incidents* describe detected malicious activities that may include a report of the techniques used, the impact of the attack, the intention, etc. *Mitigations* describe methods for resolving vulnerabilities and/or responding to incidents. Furthermore, *threats* refer to issues that must be further analysed to determine their possible implications.

Goodwin et al. provide examples such as "malware samples and stolen email addresses" for threats and argue that such information can assist towards detection and mitigation of incidents. Moving to the next layer of the pyramid, *situational awareness* describes specific types of information that can assist decision-makers to respond to incidents. Finally, *strategic analysis* is the highest level of the taxonomy and refers to the process of deep analysis of a plethora of information to determine metrics as well as projections of future risks. Figure 11 *also* touches the topic of how these information types

[6] IDS alerts can be considered both low-level data and detection indicators depending on whether they utilize a fine-tuned set of signatures or not. This is because an IDS without proper settings can generate very large amounts of alerts which include many false positives.

can be shared and how they can be generated. For instance, *vulnerabilities, incidents, mitigations* and *threats* can be publicly communicated.

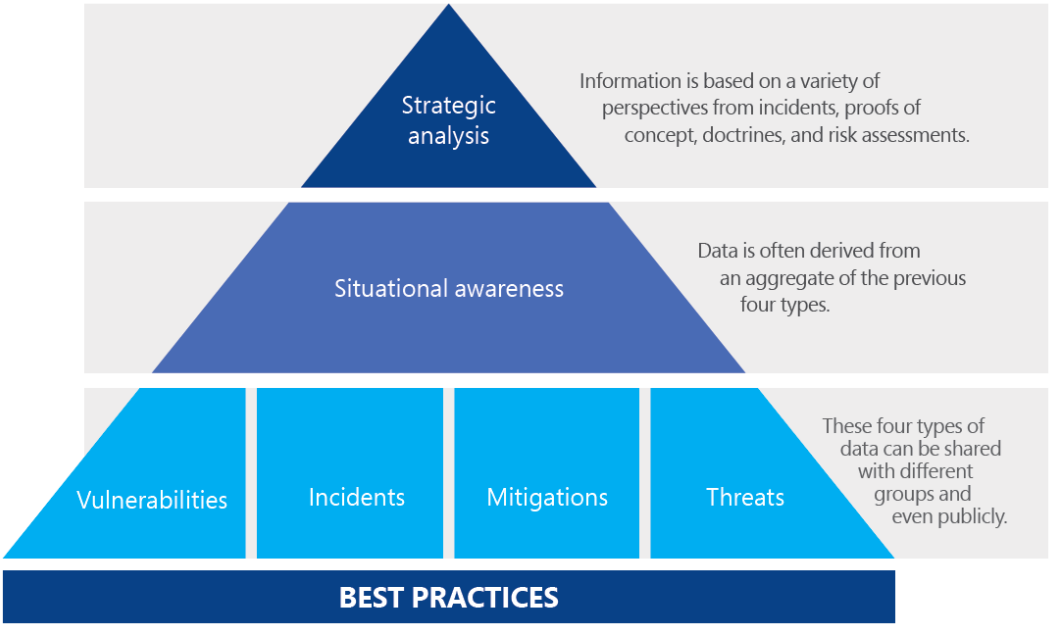

Figure 11: Types of cyber-security information. Image courtesy of Microsoft (Goodwin, et al., 2015)

2.2.4 Mechanisms of TI Exchange

In order to exchange TI among different communities, we first look at different sharing architectures and investigate their pros and cons. Then, we survey different information exchange formats and underlying methods of exchange. In this section, we also look at existing tools and platforms that facilitate the exchange of TI feeds. Effective and efficient threat information exchange requires trust relationships among the community members and data feeds of good quality.

2.2.4.1 TI Exchange Architectures

Most sharing communities exchange TI using two basic information sharing architectures: i) centralised sharing architecture; and ii) peer-to-peer sharing architecture. The two architectures are often combined to provide a third, i.e. a hybrid sharing architecture. In the following, three types of information sharing architectures are discussed with respect to their benefits and demerits.

Centralised architecture (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) is usually denoted as “hub-and-spoke”, where a central “hub” acts as a repository for information that it receives from the spokes, i.e. participating members or any other sources. Information provided to the central repository (hub) by participating members is either directly forwarded to the community members or enhanced in some way by the hub before it forwards or distributes to the designated community members. The enhancements may include aggregation and correlation of information from different sources, normalisation, enrichment of information considering additional context and quality of information. Benefits of using a centralised architecture are largely dependent on the services offered by the central hub or authority. Some of the hubs may simply broker the threat information exchange and other types of hubs may perform additional processing to enrich the threat information. Centralised hubs that use open, standard data formats and transport protocols alleviate the need for community members to adopt multiple formats and protocols to exchange threat information with others. Additionally, community members do not have to maintain many connections once the connection to the centralised hub is established. A drawback of using this architecture is that the threat information exchange is fully dependent on the central hub – this makes it a single point of

failure, can cause delays due to such as network congestion and processing backlog, and furthermore, compromise at the hub. Another drawback is that all community members are affected, if the central hub is not properly functioning or performance is not satisfactory. Finally, the centralised hub is an attractive target for attack.

In **peer-to-peer architecture** (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016), participants share information directly with each other, rather than routing information through a central repository (hub). Therefore, each of the participants takes care of enrichment processes including protecting and distributing information to the community members. The peer-to-peer architecture has a number of benefits. First, the threat information is shared in a peer-to-peer model, therefore it allows information to be distributed rapidly among each other, ii) this architecture gives more resiliency as information is available through different channels and does not represent a single point of failure or obvious target of attack. The peer-to-peer architecture has the following drawbacks: i) Peer-to-peer architectures that do not support standard formats and exchange protocols face difficulty to scale since peers must support different formats and protocols, ii) As the number of peers grow in a community, the operating costs managing connections, information, e.g. collecting, enriching, protecting, and exchanging, and trust relationship will grow exponentially.

Hybrid architectures (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) combine the advantages of both centralised and peer-to-peer architectures. In a hybrid architecture, a central hub may be responsible for resource discovery, to broker sharing requests or as a trusted third party for authentication. For example, an organisation might exchange low-level intrusion alerts using a peer-to-peer architecture but send enriched alerts or incident reports to a central hub. Another use case involves sending the same information to peers individually, as well as to the central hub. It enables both rapid action on time-sensitive data and make use of the hub's ability to gather, analyse, and correlate data from multiple community members to define long-term strategies and actions. However, a hybrid approach can increase the operating costs and make implementation difficult.

2.2.4.2 TI Exchange Types and Formats

Information is exchanged through both manual (person to person) and automated (machine to machine) ways (Goodwin, et al., 2015). Examples of manual intelligence sharing include email, phone calls, face-to-face meetings, teleconferencing etc. The complexity of cyber threats is increasing day by day which generates challenges with manual exchanges like: speed, relevance, accuracy, scalability, time etc. (Kijewski & Pawliński, 2014). In this context, automated exchange formats that offer real-time, network-speed and machine-to-machine TI exchange already exist.

In order to exchange the TI, TM Forum (TeleManagement Forum, 2013), ENISA (ENISAa, 2014) and the MACCSA (MACCSA, 2013) identify the need for a taxonomy to enable interoperable exchange of information between different implementations of open source or vendor products and consistent use of information which is likely to be published as a formatted report or as structured data. Howard and Longstaff (Howard & Longstaff, 1998) developed a common language for computer security incidents. ENISA (ENISAa, 2016) has developed a threat taxonomy that classifies threats into nine categories, namely: *Nefarious Activity/Abuse, Eavesdropping/Interception/Hijacking, Outages, Failures/Malfunctions, Damage/Loss, Disasters, Unintentional damages*, and lastly *Physical attacks* (ENISAa, 2016).

In another ENISA report (ENISAa, 2014), 53 different information sharing standards which are a mix of formats, protocols, technical approaches and frameworks in common use were compiled. Of the several standards identified (ENISA, 2013) (ENISA, 2014), *Structured Threat Information Expression*

(STIX) (Barnum, 2012)[7], *Incident Object Description Exchange Format* (IODEF) (Danyliw, Meijer, & Demchenko, 2007) and OpenIOC (Obrst, Chase, & Markeloff, 2012) were identified as the most popular in an exploratory study of 22 such platforms conducted by Sauerwein et al. (Sauerwein, Sillaber, Mussmann, & Breu, 2017). Burger et al. (Burger, Goodman, Kampanakis, & Zhu, 2014) go further and develop a taxonomy model for threat sharing technologies and using their model examine the IODEF and STIX data formats and their respective transport protocols Real-time Inter-network Defense (Moriarty, 2012) and TAXII (Connolly, Davidson, & Schmidt, 2014). Barnum (Barnum, 2012) proposed STIX/TAXII to share and structure cyber threat information.

While STIX is emerging as a popular standard, ontological (taxonomy is this context) issues is one of three main issues addressed in Serrano et al.'s (Serrano, Dandurand, & Brown, 2014) paper on the design of a cyber-security data sharing system. In order to translate between STIX and Veris[8] as an example, they propose an agile data model that can conceptually support any data model including version control of a single data model. Wagner et al. (Wagner, Dulaunoy, Wagener, & Iklody, 2016) presented MISP[9] and addressed the challenge to satisfy all users with a centralised pre-defined set of definitions with the introduction of a taxonomy based on a triple tag structure. The triple tag structures are published for reuse or new structures can be defined to meet the information sharing needs of users. ENISA (ENISA, 2016) (ENISAb, 2016) documents a number of good practices, identified by both CSIRTS and from studying the state of the art, for taxonomies.

The IDEA schema seeks to follow a number of these good practices and is described (Kacha, IDEA: Designing the Data Model for Security Event Exchange, 2013) (Kacha, IDEA: Security Event Taxonomy Mapping, 2014) (Kacha, IDEA: Classification of security events, their participants and detection probes, 2015), as simple, searchable, avoids recursion, uses unambiguous data types and semantics, and allows explicit anonymization. It has a straightforward representation, is easy to analyse by machine and supports explicit data incompleteness. Other taxonomies include *Intrusion Detection Message Exchange Format* (IDMEF) (Debar, Curry, & Feinstein, 2007) developed for exchanging information about security events between detection probes and X-ARF[10] which is based on email formatting for distributing incident reports.

2.2.4.3 TI Exchange Tools/Platforms

Automated TI exchange tools/platforms have been developed for the (automated) distribution of data collected from multiple sources and distributed to potentially impacted parties. Several organisations, including CESNET have developed TI sharing platforms. CESNET has developed an open-source efficient information sharing tool named Warden[11] (Kacha, M. Kostenec, & Kropacova, Warden 3: Internet Threat Sharing Platform, 2016). Warden handles real-time cyber threats, network speed and machine-to-machine intelligence exchange. MISP, which has been already mentioned above, is a platform that allows the collection and sharing of information on targeted attacks in a trusted environment. It uses triple tag taxonomy and trial-and-error algorithm in its data model for exchanging the information. The taxonomies and schemas are used by these automated exchange tools/platforms as reporting formats. For example, Warden uses IDEA, Email::ARF[12] and AbuseHQ[13] uses X-ARF[7],

[7] https://stixproject.github.io/
[8] http://veriscommunity.net/
[9] http://www.misp-project.org/
[10] http://x-arf.org/
[11] https://warden.cesnet.cz/en/index
[12] http://search.cpan.org/~rjbs/Email-ARF-0.010/lib/Email/ARF.pm
[13] https://www.abusix.com/

Collective Intelligence Framework[14] (CIF) uses IODEF (Danyliw, Meijer, & Demchenko, 2007), MANTIS[15], Microsoft Interflow[16] and TAXII (Connolly, Davidson, & Schmidt, 2014) use STIX (Barnum, 2012).

The standard reporting formats store and exchange actionable information in a particular data structure (ENISA, 2014). Examples of such data structure include freeform text, raw logs, character-separated values (CSV), Javascript Object Notation (JSON) etc. Warden, STIX 2.0 and MISP uses a JSON data structure and STIX 1/TAXII uses eXtensible Markup Language (XML), protobuff etc. In TI exchange tools/platforms, these data structures are used in data collection and distribution sections. For example, MISP can import data from webform, XML, OpenIOC and CSV whereas, it uses OpenIOC, IDS signatures, XML and CSV for exporting (ENISA, 2014). Warden collects data from multiple detection systems (e.g. IDSs, honeypots, firewall etc.) deployed in participating organisations and distributes it to subscribed clients using an Application Programming Interface (API) based on Hypertext Transfer Protocol (HTTP) (ENISA, 2014).

2.2.4.4 TI Sharing Quality

In order to facilitate effective and efficient sharing of TI feeds within the community of NRENs, one has to make sure that TI feeds are of good quality. The quality of decision-making depends on the quality of available TI feeds (MACCSA, 2013). According to (Mohaisen, Al-Ibrahim, Kamhoua, Kwiat, & Njilla, 2017) (Habib, Volk, Hauke, & Mühlhäuser, 2015), without high quality of shared information, no actionable intelligence can be obtained. As the quantity of intelligence rapidly grows due to an increasing number of incoming TI feeds, human operators become the bottleneck for assessing the quality of these feeds. Automating the quality assessment of TI feeds remains a challenging task and is tied to the question of establishing trust on the feeds.

The quality indicators or aspects are of paramount importance in the context of establishing trust on the TI feeds: a timely indicator, like a source of the attack, can be used to defend against an emerging attack. According to (Mohaisen, Al-Ibrahim, Kamhoua, Kwiat, & Njilla, 2017), one way to deal with the quality of indicators is to use historical or prior information provided by various sensors or community members as a metric for their quality. However, such an approach has some shortcomings. It considers the sources of indicators, but there could be multiple aspects influencing each of the indicators. Certain community members and sensors might be well-known for certain indicators, e.g. domain names and accuracy of TI feeds. Combining multiple of these indicators in assessing the TI quality score is considered much more beneficial for the community members than providing a single score (ENISAa, 2014) (Mohaisen, Al-Ibrahim, Kamhoua, Kwiat, & Njilla, 2017). Recent advances in computational trust models as well as machine learning techniques and their applications to security scenarios (Habib, Volk, Hauke, & Mühlhäuser, 2015) (Jang, Kang, Woo, Mohaisen, & Kim, 2015) could be fruitful direction to achieve multi-indicator or multi-dimensional trust scores of TI feeds.

2.2.4.5 Privacy-Preserving Sharing

With Organisation-to-Organisation interactions (both public and private sector) both personal and commercially relevant data is likely to be present. The decision to allow and disallow usage of data is normally not in the hands of a single individuals, but potentially large organisations. The data in question may be tied to other preferences, legal contracts and national laws, meaning the requirements for the PROTECTIVE users are different from that of personal users.

[14] http://csirtgadgets.org/collective-intelligence-framework/
[15] http://django-mantis.readthedocs.io/en/latest/readme.html
[16] https://technet.microsoft.com/en-us/library/dn750892.aspx

A static compliance-checking framework showing that executing business processes satisfy certain specifications was described by Liu et al. (Liu, Muller, & Xu, 2007). Their approach involves transformation of *Business Process Execution Language* (BPEL) models to pi-calculus processes. The focus here (again) is not on monitoring actual data-flow compliance as we concern ourselves, but on business process compliance.

Compliance monitoring observes business process execution and reports violations of specific laws, regulations or contracts. Several commercial products exist that focus on compliance with information security regulation and standards such as ISO 27001 (Calder & Watkins, 2008) and Sarbens-Oxley (Sarbanes, 2002). However, to our knowledge there are no equivalent products focused on TI sharing data handling.

The EU project Centre on Migration, Policy and Society (COMPAS) has developed a business compliance framework for *Service-Oriented Architectures* (SOAs) (Daniel, et al., 2009). COMPAS uses finite state machines to recognise patterns of events; these events are at system-level that have relevance for compliance. A pattern identifies how the system violates a specific compliance requirement. The state machines report the occurrence of patterns using a corresponding high-level event. Monitors designed using the COMPAS approach aim to check compliance to particular business processes using run-time access to models of correct behaviour. In contrast, we seek monitors that check for satisfaction of compliance criteria. Our approach also uses finite state machines.

Soto-Mendoza et al. (Soto-Mendoza, Serrano-Alvarado, Desmontils, & Garcia-Macias, 2015) proposed an approach to derive privacy policies based on semantic web technologies. Their composition of rules are based on the data-usage context and deduces implicit terms, including uses basic operators and ontology-based rules to consider data-usage context. It is worth noting the authors point out that inconsistencies exist, and that these can be minimised with contextual rules that incorporate priorities.

2.3 Case Studies

We present several case studies to illustrate existing practices in TI sharing, as described by the literature and PROTECTIVE's NRENs. This aims to help inform us about the requirements and specifications in Section 3.

In cybersecurity, every information sharing platform demands guiding principles to be followed for development and operation of information sharing in a secure system. Sedenberg et al. (Sedenberg & Mulligan, 2015) identified in the public health domain that the existing information sharing proposals do not properly addresses the kinds of information to be shared and policies required to understand the requirements of a contemporary information sharing environment. In the public health domain, information is collected and shared to examine public health vulnerabilities within the population and particular communities. Also, the public health activities monitor the disease to identify and eradicate its root cause. However, information is shared in a complex environment for protecting privacy while requiring maximal participation.

If information is not shared carefully then it may lead to social stigma and discrimination which may stop an individual from seeking proper care. The *Centers for Disease Control and prevention* (CDC) has prepared guidelines using 18 variables considered identifiers, under the *Heath Insurance Portability and Accountability Act* (HIPAA), which must be removed from dataset before sharing in order to ensure confidentiality. CDC also provides a Certificate of Confidentiality for authorising researchers to protect the privacy of individuals so that no federal, state, local, criminal, administrative, legislative, or any other proceedings can compel the release of identifying information without the individual consents.

Lewis et al. (Lewis, Louviens, Abbott, Clewley, & Jones, 2014) proposed a different approach in his framework for information security sharing for SME supply chains. This work analyses the implications of adopting cybersecurity metrics for information sharing in 17 UK SMEs and it is analysed that there are two major issues in implementation of cybersecurity information sharing schemes: complexity of the legal guidelines (Aviram & Tor, 2004) and risks associated with sensitive security breaches when sharing information (Boyens, Paulsen, Moorthy, Bartol, & Shankles, 2014).

To address complexity of the legal guidelines, *Information Sharing Agreements* (ISA) or *Cyber Information Sharing Agreements* (CISAs) amongst SMEs are proposed on which business may be conducted. It has been found that the types of information that could be shared include: speed of patching vulnerabilities, total number of attacks and properly configured resources. Risks exposure of SMEs in sharing information include: negative exposure or misuse of data, loss of reputation, poor quality of information and exposure of organisation's weaknesses in cyber defence capability. A number of these problems still have to be resolved.

Kaijankoski (Kaijankoski, 2015) examines the banking and finance sector to evaluate the effectiveness of public-private partnerships to advance cyber information sharing. This work confirms that private sector companies do not show interest in sharing threat information due to lack of clear guidelines (Peretti, 2014) and sensitivity of sharing information. Lack of clear guidelines abide the organisations to follow the existing standards which deter many organisations from collaborating. Private organisations have the fear of government security leaks; thus these organisations avoid access to sensitive information. However, recommendations are made to integrate the information at some central trust point but these recommendations lack in real plan of action or policies.

Arizona Cyber Threat Response Alliance (ACTRA) (Haass, Ahn, & Grimmelmann, 2015) (ACTRA, 2012) is a non-profit organisation developed with joint venture of FBI's InfraGard and the *Arizona Counter Terrorism Intelligence Center* (ACTIC) for improving security. This case study discusses the issues, investigations and solutions, as well as recommendations for newly forming information sharing groups in organisational, technical and legal domains. It indicates that there is need of structured formats, systematic procedures, framework and cultural development in order to share relevant attacks details. In order to secure a community, it also states the importance of standardising the intelligence formats for machine-to-machine and other sharing purposes. A survey was conducted to understand what kinds of information would be valuable, who the expected audience was and how the data should be delivered. This work proposed six recommendations for removing the existing information sharing barriers, and further concluded that the hindrances of sharing information among multi-sector will disappear with establishment trust and improvement in technology and policies.

NIST (Johnson, Badger, Waltermire, Snyder, & Skorupka, 2016) presented scenarios that describe TI sharing in their "*guide to cyber threat information sharing*". The scenarios demonstrate how sharing and coordination can increase the efficiency and effectiveness of an organisation's cybersecurity capabilities. The authors highlight that the scenarios only represent a small number of possible applications of information sharing and collaboration. The cases listed include:

1. **Nation-State Attacks against a Specific Industry Sector** – discusses phishing attacks in an attempt to extract private data. As soon as one company's security team identifies a new attack, it shares data with its peers within a forum. In the example, they illustrate how Company A in the forum has advanced malware analysis capabilities and is given a malware sample by Company B. Company A then shares back the insight gained through its analysis of the malware provided by Company B. The community as a whole benefits from the synergy between Companies A and B.

2. **Campaign Analysis –** outlines how different companies can share independent analysis on the same threat challenges. This allows analysts to discover patterns across companies, and while sharing intelligence it may be possible identify whether there are any larger coordinated threats.
3. **Distributed Denial-Of-Service (DDoS) Attack against an Industry Sector –** describes the benefits of TI sharing relationships between ISPs, law enforcement, legal bodies and other companies in the face of hacktivist groups conducting coordinated resource exhaustive (DDoS) attacks. This point is related to the first scenario – the key difference being that different capabilities across different bodies serve different purposes. For instance, an ISP can aid in load balancing in active attacks, or law enforcement to aid in criminal investigations. This is a different from 1) in that the TI sharing is not related to enhancing insight, but the TI sharing aids actions.
4. **Financial Conference Phishing Attack –** highlights the uses of conference calls (or other forms of informal channels) to share information to share threat indicators to non-technical audiences.
5. **Business Partner Compromise –** discusses how security teams of two companies had agreements and processes in place for a joint response, and having pre-established contacts and existing trust relationships, and had already understood each other's networks and operations, the companies were able to quickly respond and recover from the incident.
6. **US-CERT Provides Indicators, Receives Feedback –** describes that while investigating incidents, affected companies may also be able to identify new indicators or provide context regarding the attack to a CERT (e.g. US-CERT). The CERT is then able to share these new indicators with other firms after anonymizing the sources, leading to a more comprehensive response to the threat.
7. **A Retailer Fails to Share –** discusses the consequences of failing to share TI in the retail sector. The scenario provides insight into how attackers can reuse attacks across different companies, but also the losses involved (esp. financial losses).

3 Requirements and Specifications for TI Sharing

3.1 Requirements

This section describes the overall requirements and specifications based on the background literature, questionnaire, interviews and observations to date[17]. Some of the high-level requirements for the system include (the IDentifications (IDs) referenced here can be found in D2.1):

- **Deliver a common understanding of what TI encompasses** – presently, different organisations use different definitions. PROTECTIVE will use the definition used by ENISA as discussed earlier. This is a non-functional requirement.
- **Ability to share TI –** enabling PROTECTIVE partners to set up communities, share automated and manual TI straightforwardly. See **IDs: IF-01-10**.
- **Ability to compute trust factors** – enabling analysts to make better use of TI they receive by understanding how much confidence they can attribute to the TI they have received. This is covered in **ID: TR-01**: *The reputation of a "foreign" IP may be used to prioritise which information needs to be processed first*.
- **Ability to apply restrictions in TI sharing** – limiting TI sharing to adhere to legal frameworks, organisation preferences and non-disclosure agreements. See: **ID: IF-09**: *configure exactly what information may be disclosed to other participants, to follow the GDPR and internal policies.*
- **Define appropriate sharing etiquettes, procedures/protocols for PROTECTIVE communities** – the means to cement best-practices have not been established, and will be continually refined (see D2.4 for an in-depth discussion on this issue). This includes understanding communities in the context of TI, while it is necessary to understand how to address technical requirements for TI sharing, human-factors of what makes up a community and how to encourage appropriate sharing also needs to be considered. This is a non-functional requirement.

3.2 PROTECTIVE TI Sharing Features

PROTECTIVE aims to facilitate TI sharing among the NRENs and SMEs. The features include configuring and managing community members, providing an easy-to-use schema for different TI types, ensuring the quality of TI as well as compliance to data protection regulations while sharing the TI.

3.2.1 TI Community Management

PROTECTIVE will provide support for **private TI-sharing communities** with differing degree of formality. Within the scope of the project we envision the following types of community.

3.2.1.1 NREN Local Community

This is the NREN constituency - the country internal educational institutions and other organisations that the NREN supports. The interaction and degree of sharing between NREN and their constituents differs from country to country. For automated TI sharing, PROTECTIVE will support the default Warden/Mentat sharing model. This is described in Figure 12. Each of these constituency communities have their own participating member (e.g. member 1, member 2 and member 3) connected to a Warden module. In each of these member organisations, probes such as honeypots (Kippo, Dionaea and LaBrea) are registered as Warden sending clients which feed data into a Warden Server. Protection systems (clamav, RTBH, firewall) are connected as Warden receiving only clients. Mentat is connected to the Warden server as a sending client that also receives data.

[17] Full system requirements discussed in-depth in D2.1. For the purposes of this document, we summarise the TI sharing specific aspects of these requirements, including non-functional requirements.

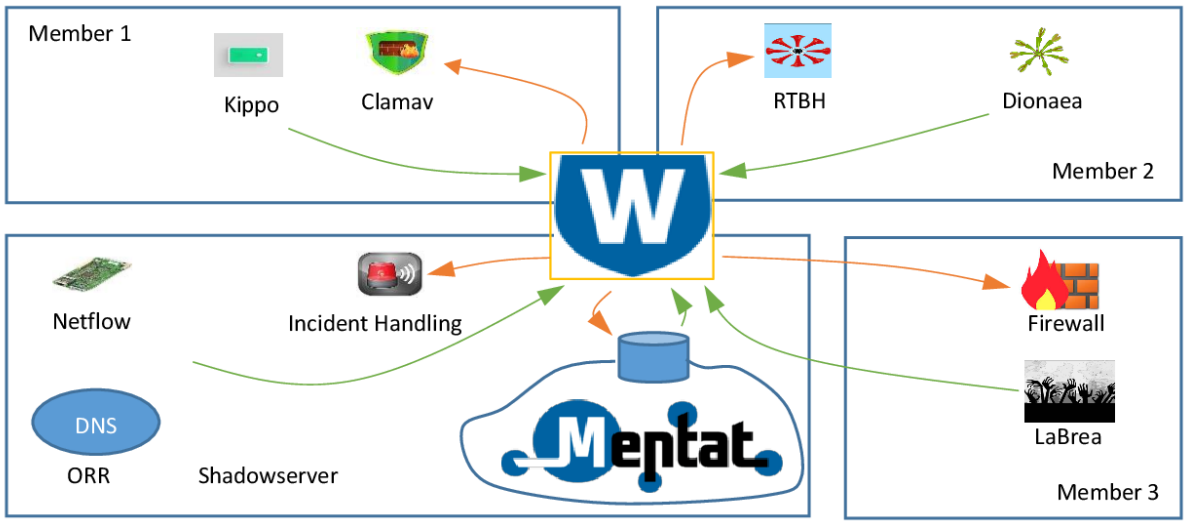

Figure 12: Constituency community architecture within PROTECTIVE

Authentication and authorisation in the Warden system: In the Warden system, two types of authentication are required: client and server authentications. A client and server authentications are processed through client and server authority certificates respectively during transition phase. Figure 13 shows the client authentication process. Initially, every client has to register with the Warden server.

After registering with the server, a certificate containing client and machine identifiers are provided to client over a secure channel. Thereafter, whenever a Warden client queries the Warden server it must authenticate using its certificate. The certificate is verified at the server side and this process is known as warden client authentication. Similarly, the client verifies the server's certificate to verify the identity of the server. After both server and client are authenticated to each other, a response to the query is generated. Figure 13 and Figure 14 demonstrates the client and server authentication process consecutively.

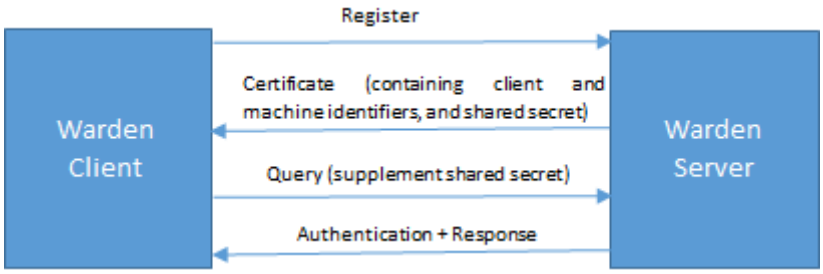

Figure 13: Client authentication process

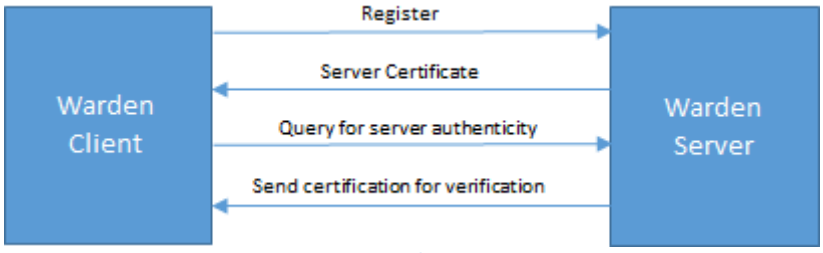

Figure 14: Server authentication process

3.2.1.2 Inter NREN Community

The inter-NREN community will consist of the project members PSNC, RoEduNet and CESNET. They will be joined by SME member TheEmailLaundry (EML). It is expected that other parties may also participate in this community over time. The organisational trust model is *Direct Historical Trust* and the information sharing model is *Voluntary*. Within the scope of the project the community formed by the partners above is expected to be somewhat more formal than the constituency sharing. The grounds and basis for information sharing will be defined by an IEP though this is expected to be a minimal set of conditions.

More generally expanded sharing with other NREN's than the consortium members will require a more formal sharing framework. However, in these communities the information sender has full control over what information is shared with other members.

The two PROTECTIVE models presented here are similar to MISP communities:

- **organisation only**: Only members of an organisation are allowed to see an event. This model is easy to implement in PROTECTIVE though it is not widely used.
- **community only**: this corresponds to the constituency community level in Protective. Users of the WARDEN community can see the event, including organisations that run WARDEN servers that synchronize with that server.
- **connected communities -** this model corresponds to the Inter-NREN case.

3.2.1.3 PROTECTIVE Community Architectural Approaches

Different architectural approaches can be used to realise the inter-NREN sharing. The choice of architecture can influence the organisational trust model. For PROTECTIVE, two architectural approaches exist: peer-to-peer and central hub. The peer-to-peer architecture will be implemented using the baseline modules Warden and Mentat as shown in Figure 15.

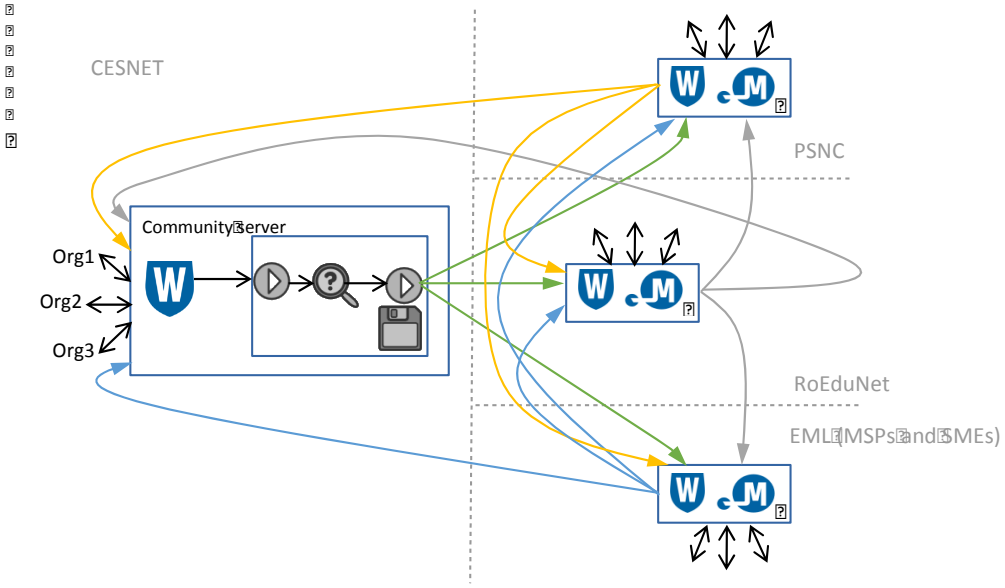

Figure 15: Peer-to-peer TI sharing architecture for PROTECTIVE Community

Peer-to-Peer Inter-NREN Architecture: In this architecture there are multiple NREN based communities (CESNET, PSNC, RoEduNet and EML) and each NREN shares information directly with every other NREN. The community server utilizes Warden and Mentat as underlying technology but the community-related components must be developed and implemented into Mentat to support selective peer-to-peer exchange. The advantages are that an NREN has full control over information

sent to other NREN s and as a typical peer-to-peer system it is more robust to potential Denial-Of-Service (DoS) attacks.

Inter-NREN Central Hub Architecture: In this central-hub community architecture information is shared via a centralised Warden hub. In this case also each NREN decides what it is prepared to share. However, this architecture entrusts the administration and operation of the sharing to a single party.

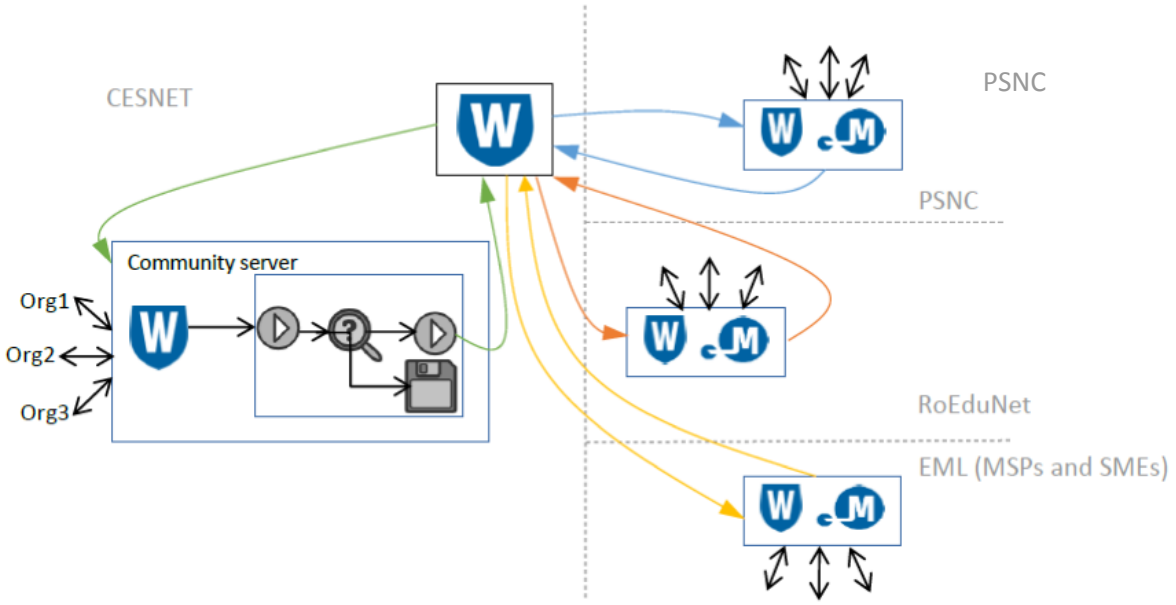

Figure 16: Central hub architecture for PROTECTIVE

This structure supports and may require a *Mediated Trust* model i.e. a mutually trusted third party organisation provides assurances that lead to a trust level between two organisations. In other words, the central hub may be operated by a trusted third party, which then arbitrates the distribution of information. This mode of operation is similar to the MISP NATO connected communities shown in Figure 9.

3.2.2 TI Information Types

The various prominent taxonomies that have been proposed in the literature have been already described in Section 2.2.3. Based on a qualitative analysis of the state of the art as well as empirical knowledge (e.g. as a result of internal PROTECTIVE communication and questionnaires with NRENs) the project will be utilizing the ENISA (ENISAa, 2014) taxonomy as a basis for its operations. In particular, in the context of PROTECTIVE the ENISA taxonomy will be used by focusing on *low-level data, detection indicators* and *advisories.* A number of non-exhaustive utilisation examples for each of the three information types can be found on Table 2.

Information Types	Utilisation (Non-exhaustive) Examples
Low-Level Data	• Raw IDS alert data (e.g., from SNORT (Roesch, 1999), Bro (Paxson, 1999), Suricata[18], and other IDSs) • Flow data (e.g., Flowmon) captured by firewalls and routers • Application-level log data (e.g., server log files)

[18] https://suricata-ids.org

Detection Indicators	• Honeypot alert data (e.g., from Dionaea, Kippo/Cowrie, HosTaGe (Vasilomanolakis, et al., 2013), etc.) • Hashes of malicious entities (e.g., from malware captured from the Dionaea honeypot) • Malicious IP addresses (e.g., from infected machines as detected by analysing raw IDS alert data and/or combining knowledge from honeypots and other sources)
Advisories	• High-level alert data as a result of sophisticated correlation and aggregation of both low-level data and detection indicators • Global statistics over a period of time as an indicator for adversarial trends and techniques • Vulnerability advisories such as automated and/or manually written technical reports that describe in detail the lifecycle of an attack

Table 2: Examples of Information types' utilisation in PROTECTIVE

From the plethora of existing alert data formats and schemas, PROTECTIVE has decided to adopt the Intrusion Detection Extensible Alert (IDEA) format (Kacha, 2013). The schema has been already introduced in the previous chapter (see Section 2.2.3) along with other approaches from the state of the art. There are two core points and arguments for the utilisation of IDEA: *simplicity* and *applicability*, both of which will be further explained in the following.

First, the IDEA schema was designed in a simple yet flexible manner by researchers and network administrators who were intending to practically utilize the format to their daily operations (Kacha, 2013). Hence, in contrast to other schemes, IDEA does not create overhead due to its complexity (such as for instance STIX (Barnum, 2012) or IDMEF (Debar, Curry, & Feinstein, 2007)), neither bounds the operations of the system due to oversimplification (Vasilomanolakis, Karuppayah, Kikiras, & Mühlhäuser, 2015). Second, the IDEA format fulfils the applicability requirement, in the sense that it has been used by NRENs (e.g., CESNET[19]) for a number of years and had practically proved its usability and flexibility. In the following figures, we provide examples that map IDEA with the two possible information types.

In more details, Listing 1[20], depicts the IDEA representation of a low-level data example. Similarly, Listing 2[20] shows a case of a detection indicator.

[19] https://www.cesnet.cz
[20] https://idea.cesnet.cz/en/examples

```
{
   "Format": "IDEA0",
   "ID": "3ad275e3-559a-45c0-8299-6807148ce157",
   "DetectTime": "2014-03-22T10:12:56Z",
   "Category": ["Recon.Scanning"],
   "ConnCount": 633,
   "Description": "Ping scan",
   "Source": [
      {
         "IP4": ["93.184.216.119"],
         "Proto": ["icmp"]
      }
   ],
   "Target": [
      {
         "Proto": ["icmp"],
         "IP4": ["93.184.216.0/24"],
         "Anonymised": true
      }
   ]
}
```


Listing 1: Low-level data and IDEA: scanning detection example


```
{
   "Format": "IDEA0",
   "ID": "2E4A3926-B1B9-41E3-89AE-B6B474EB0A54",
   "DetectTime": "2014-03-22T10:12:31Z",
   "Category": ["Recon.Scanning"],
   "ConnCount": 633,
   "Description": "EPMAPPER exploitation attempt",
   "Ref": ["cve:CVE-2003-0605"],
   "Source": [
      {
         "IP4": ["93.184.216.119"],
         "Proto": ["tcp", "epmap"],
         "Port": [24508]
      }
   ],
   "Target": [
      {
         "Proto": ["tcp", "epmap"],
         "Port": [135]
      }
   ]
}
```


Listing 2: Detection Indicator and IDEA: honeypot data example

3.2.3 TI Sharing Quality

The trust mechanism aims to improve the management, sharing, and prioritisation of threat intelligence within the community of NRENs and SMEs by determining the quality (or reputation) of TI feeds, i.e. how “good” is the feed itself. Additionally, another aim to assess the quality of a particular malicious entity, i.e. how “bad” is an entity (e.g. IP address) associated with a feed. The first variant aims to improve the management and sharing of TI feeds with respect to their quality (“goodness”) and the second variant seeks to advance the prioritisation of TI feeds with respect to their level of maliciousness (“badness”).

In PROTECTIVE’s system, we will use CertainTrust (Ries, Habib, Mühlhäuser, & Varadharajan, 2011) (Habib, Ries, Hauke, & Mühlhäuser, 2012) (Habib, Volk, Hauke, & Mühlhäuser, 2015) computational trust methods to assess the quality of TI feeds. In some cases, the quality (reputation) assessment becomes uncertain due to incomplete and insufficient data. CertainTrust provides a mechanism to quantify the level of confidence (certainty) along with the quality of TI feeds. The quality assessment

of the TI feeds are based on multiple aspects or dimensions such as accuracy and timeliness of the feeds. CertainTrust method has the ability to compute and visualise quality of the TI feeds and associated confidence level based on multiple aspects. NREN CSIRT analyst should have the capability of providing feedback on the overall quality and individual quality aspects. We aim to develop a usable feedback mechanism as part of the visualisation mechanisms leveraging the CertainTrust method.

3.2.4 TI-Sharing Compliance

PROTECTIVE has set out a wide-range system for privacy and ethics governance. It is necessary to consider how such governance will affect sharing capabilities and capacities. As discussed in-depth in D2.4, we will comply with the national and EU ethical and legal framework, such as the Charter of Fundamental Rights of the European Union, the European Convention on Human Rights, the European Code of Conduct for Research Integrity, and the Horizon 2020 Rules for Participation once the PROTECTIVE tool itself begins to share TI with organisations.

Our approach for legal and regulation compliance is a six-layered mechanism that sets out to address privacy and ethics concerns during TI sharing in the tool itself[21]. These layers include: 1) an external advisory board; 2) an ethics committee at academic institutions; 3) the Trusted Introducer Service[22]; 4) the local Data Protection Authority (DPA); 5) A set of (human-level) protocols among PROTECTIVE partners; and 6) making use of a compliance module[23]: a monitor and enforcer consisting of compliance rules added as a last step to check that data leaving an organisation does not violate GDPR or NDA-related concerns. Its design will help us to review data in-transit between NRENs and allow us to implement mechanisms that not only ensure data sharing does not go beyond explicit rules of sharing in spirit, but also (demonstrably) to the letter.

In 6), we are interested in detecting violations of the various preferences expressed including GDPR, NDAs, organisation-specific rules and checking for well-formed data. This need goes beyond the legal requirements and makes it necessary to consider establishment of a basic formal description specifies checks of restrictions and conditions from a regulatory and legislative level down to the technical level. This is a TI-sharing checker that will exist at the edge of an NREN that reads all TI leaving the organisation and enforces changes to be made in the TI if it is in violation of any rules, if any TI is about to leave the premises by mistake or in incorrect form, the compliance module's task is to:

- **conduct** exception handling of corrupt or missing TI
- **detect** any TI sharing violations - i.e. the information being sent out does not conform to the rules specified by the module. These rules are specified by the GDPR and NDAs in place at the NREN, maintained by a PROTECTIVE partner.
- **prevent** sharing violations from happening after detection - either by *dropping* the sending of the TI entirely or *correcting for* it.
- **log** the incident so this type of violation can be corrected for and does not occur again (or at least drastically minimised from happening again), but also provide empirical evidence that PROTECTIVE is making its best effort to comply with the GDPR and NDAs.

The approach also needs to be able to identify which data attributes are classified as sensitive data[24] and not, but also guidelines for which to derive other personal data in the future. Such a compliance

[21] More on this in D2.4.

[22] https://www.trusted-introducer.org More on this in D2.4.

[23] Use case 6 in D2.4 describes this module in depth. Further discussions about the sharing model behind the compliance module can be found in D5.1.

[24] Note: in this context we make a separation between personal and sensitive data. Personal data *"shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is*

monitor would include: a data protection policy, a policy on retention periods for all items of sensitive data, procedures for automatically handle access requests from individuals, guidelines to ensure staff appropriately trained in data protection, and a plan for regular reviews and audits of the data held and the manner in which the data is processed.

3.3 PROTECTIVE TI-Sharing System

PROTECTIVE will develop a computing platform that will provide the cyber-situational awareness as well as developing the policies and mechanisms to enable threat intelligence sharing between CSIRT teams in a grander PROTECTIVE TI community or ecosystem. Such an ecosystem is a federation of a number of PROTECTIVE nodes in different partner organisations which share information for the purposes of mutually improving identification and prevention and mitigation of threat events in their respective constituencies.

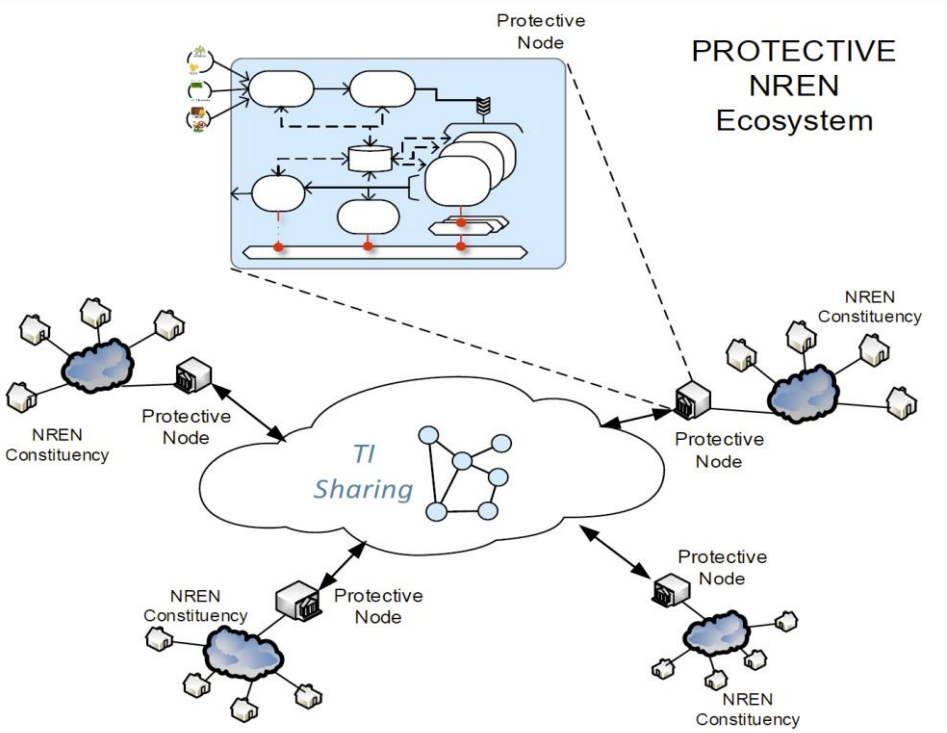

Figure 17: The PROTECTIVE System

Figure 17 illustrates the ecosystem of PROTECTIVE. This shows a number of NREN networks with their constituency members. Each network has (at least) one PROTECTIVE node which is used to fulfil the Cyber Security Awareness (CSA) goals above, and also to route threat information to and from community partner PROTECTIVE node. This is depicted by the double ended arrows. The figure also shows a detailed snapshot of a PROTECTIVE node to expose some of the key components of the node (which will be discussed below).

one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" as specified by the European Parliament and Council in 1995 (please refer to D2.4 for a more in-depth discussion on this). Sensitive data simply means content that must be protected from unauthorised access to safeguard the privacy, safety or security of an individual, organisation, asset, service or infrastructure.

3.3.1 TI Sharing Architecture

PROTECTIVE TI sharing includes the sharing of detection indicators via IDEA and the sharing of reports and advisories via email. A number of PROTECTIVE node functions are involved in the complete TI sharing i.e. within an NREN constituency as well as with external partners including other NRENs. These functions include – see Figure 18 below:

- TI Sharing Subsystem;
- TI Trust component in the Enrichment subsystem; and
- Reporter subsystem.

Each of these is described in more detail in the following sections.

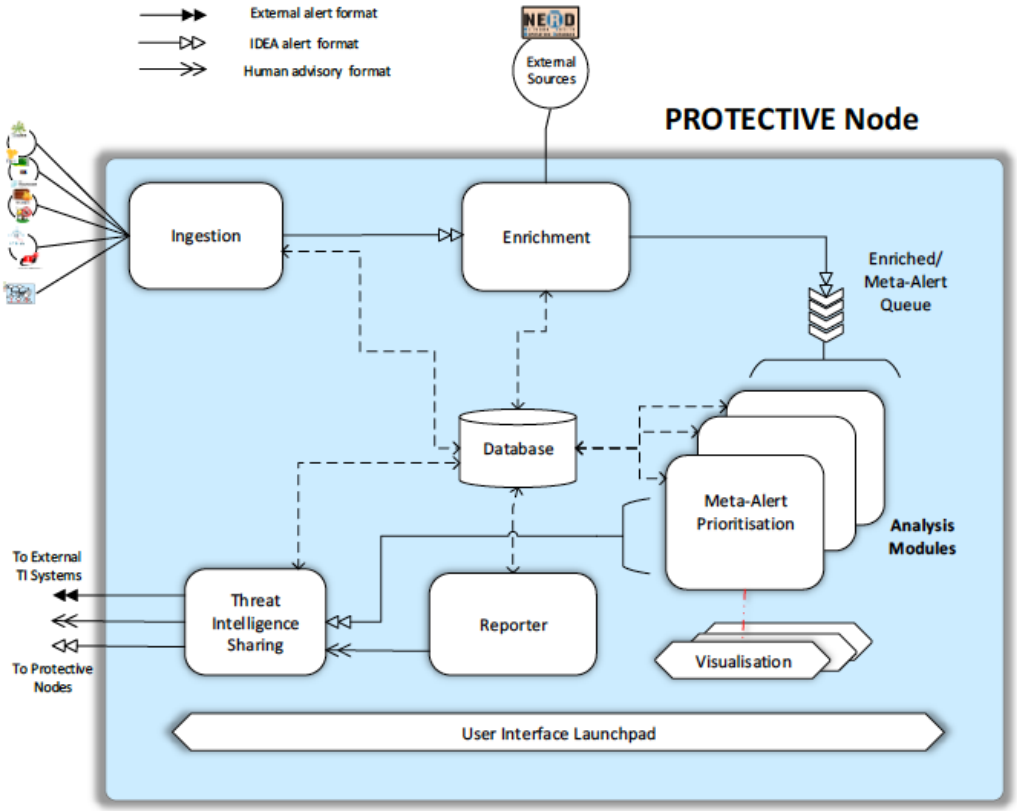

Figure 18: PROTECTIVE Node (constituency)

3.3.2 TI Sharing Subsystem

The TI sharing subsystem implements the function of TI Distribution and TI Admin. In Figure 19, the double headed open arrow represents the flow of TI emails. These originate from the Reporter function and are routed through the **Compliance Manager** for checking and subsequently dispatched. The double headed closed empty arrow represents IDEA flows. This TI also flows through the Compliance Manager and subsequently is routed through the **TI Router** (Warden). The closed black arrows represent the case where the IDEA format is converted to an external format e.g. STIX for distribution.

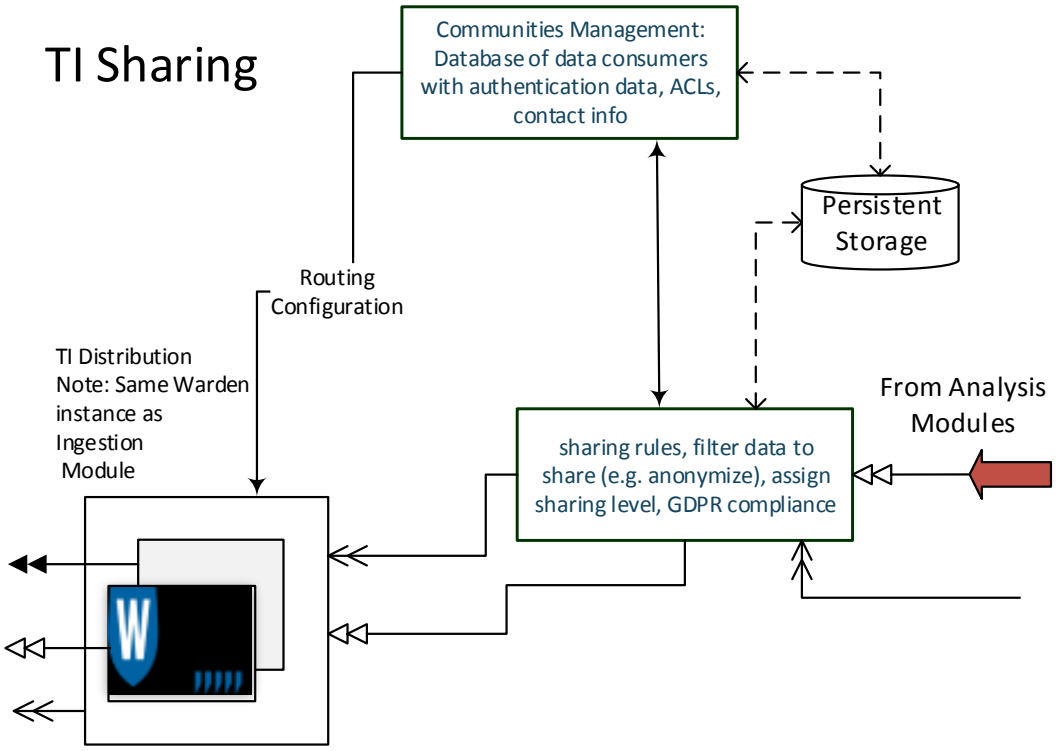

Figure 19: PROTECTIVE TI Sharing Subsystem

The TI Administration concerns primarily the **Communities Manager** function in Figure 19. The exact scope of this function will depend on the TI community model chosen. The general functions included in this function are:

- User Management;
- User Authentication and Authorisation;
- TI Distribution channel authentication and management; and
- Configuration of TI routing.

User management entails the creation and deletion of user and related activities. In the case of most community models a TI sharing partner will control its own PROTECTIVE sharing node and the users will be members of its own organisation. User management in this case will follow standard "admin/user" roles. In some cases, an NREN may share access to received events with its constituency members and in this case selective exposure of stored TI may apply and corresponding policy based authorisation mechanism needed. In the case where a trusted third-party sharing model applies a more formal sharing arrangement will be required and this will require, in turn, a comprehensive authorisation and policy management system, with several levels of role. Authentication in all of these cases will depend on the IEP – in most cases it will use a conventional userid/password, and in some cases, requiring two-factor authentication.

TI Distribution consists of the Compliance Manager function and the TI Routing function. The compliance manager applies a set of rules and policies to TI information flows to ensure that they comply with the prescribed IEP policies including GDPR. This may also include pseudonymization functions. NREN's willing to share information with other NREN's will register to the Warden of the receiving NREN. The registration procedure will take place over a trusted and encrypted channel. TI distribution channels will then be secured using Hypertext Transfer Protocol Secure (HTTPS) and X509 keys. TI routing determines what information gets sent to whom. This will be configured centrally for each PROTECTIVE sharing node and routing tables or rules will be distributed to the TI distribution

nodes i.e. the Compliance Manager and the TI router (Warden). Routing information is updated from the Communities Manager.

3.3.3 TI Trust Component

TI Trust component is part of the TI Enrichment subsystem. Aggregated and normalised alerts are enriched with quality scores and the reputation scores of the entities responsible for the alerts. The TI Trust component deals with quality (reputation) assessment in two contexts: one is to determine the quality of threat intelligence feeds and the other is to ascertain the reputation of malicious entities associated with feeds. Enriched alerts are later ranked and prioritised.

The TI Trust Component includes two Trust Specification modules, two Machine Learning modules and a common Trust Computation module. Figure 20 provides a generic architecture of the module.

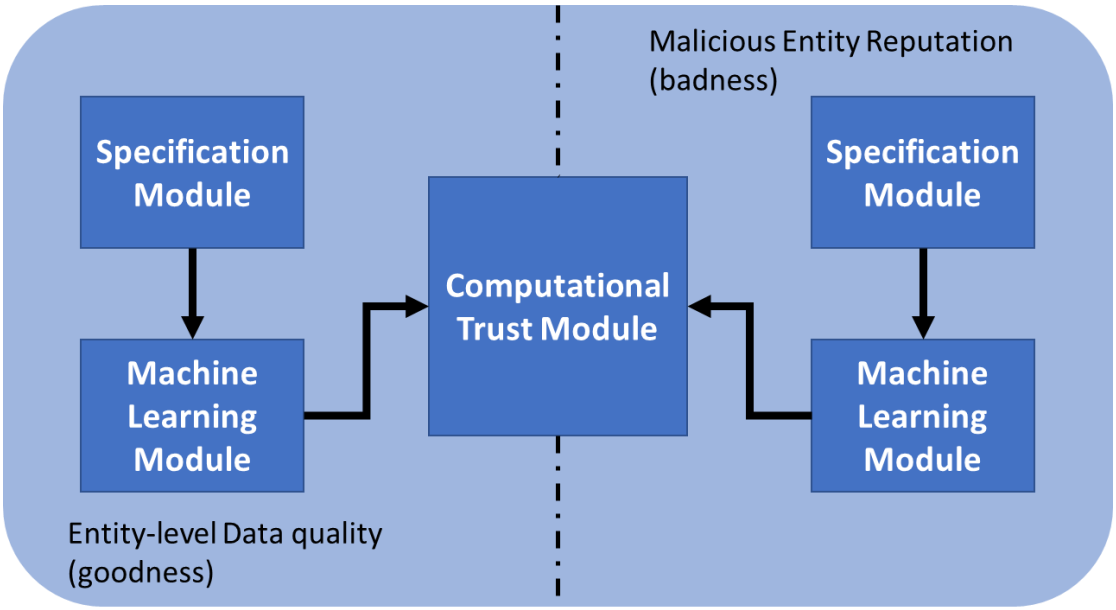

Figure 20: TI Trust Component

3.3.3.1 Trust Specification Modules

These modules specify the properties according to the context of reputation calculation. In order to calculate the reputation of TI feeds, the module will specify the following properties: accuracy, completeness, timeliness, freshness, relevance, reputation of the feed distributor (i.e. community member) (ENISAa, 2014). In order to calculate the reputation of malicious entity, the module specify the following properties: type of sensor/detector, timestamp, quantity of alerts in a certain time-window, variety of alerts in a certain time-window, variety of sensors that detected the entity (IP address) as malicious, evidence from external sources, AS ranking, and static vs. dynamic IP address.

3.3.3.2 Machine Learning Modules

The Trust Component contains two machine learning (ML) modules for two contexts. In the context of data quality, ML module will be used to build stereotypes of community members, i.e. in possession of similar certifications or belong to same domain of business. Stereotyping (Burne, Norman, & Sycara, 2013) is used to assign an initial trust level (score) to a new member in a community. This trust score is an input parameter of Computational Trust Module. The properties (or features) from the corresponding Specification Module will be served as inputs to the ML module. In the context of malicious entity reputation, ML module will be leveraged to estimate the reputation of malicious entities. In order to estimate the reputation score, machine learning mechanisms, e.g. Deep Learning and Long Short Term Memory (LSTMs) (Yann, Bengio, & Hinton, 2015), will be used on the features (properties) specified by the corresponding Specification Module.

3.3.3.3 Trust Computation Module

This module contains computational trust methods, CertainTrust to compute the reputation scores. CertainTrust Software Development Kit (SDK) will be instantiated and extended to deal with different types of properties. The SDK provides various functions to aggregate multiple properties to generate an overall reputation score. These functions will be leveraged to generate quality (reputation) of the TI feeds as well as reputation scores of the malicious entities. The SDK also provides visualisation function to communicate multi-dimensional reputation score to security analysts in NRENs and SMEs. Security analysts can use the visualisation to provide feedback on the data quality or reputation scores.

3.3.4 Reporter Subsystem

The **Reporter** subsystem is intended to share "Advisory" type TI (as per the categorisation described earlier) with PROTECTIVE community members. It is based on the existing Reporter function in Mentat. This TI will be based both on security alerts generated within the PROTECTIVE system and also on TI received from sources external to PROTECTIVE.

The reports are asynchronous to the alert data stream i.e. they are generated periodically from information stored in the database according to a schedule defined by the administrator. The reports are generated, therefore, from data aggregated over the reporting period. Reports are customised for each receiver i.e. a community member may receive only information that is relevant for its own IP address range – particularly appropriate for NREN constituency members – and/or according to the type of information content of the alert.

Reports are used to inform community members of vulnerabilities or threats relevant to their community members. Reports could, for example:

- identify operational and security problems (overload circuits, poor configuration of firewalls or servers enabling misuse attacks, DoS / DDoS, etc.). The observed data may allow removal of such problems before they can be exploited by attackers;
- detect attempts to exploit infrastructure and data; and
- indicate threats that are relevant for a particular asset configuration e.g. critical vulnerabilities for particular OS versions.

This latter report entails interaction between the Reporting subsystem and the Context Awareness subsystem – see D4.1 – to correlate information from both sources. Parsing information from external advisories originally intended for human consumption may require a machine learning solution if correlation of such reports with the alert database is required.

Reports can be classified in different levels to severity, for instance, according to the following criteria:

- Low risk - Information / no response needed
  - few critical network traffic
  - spam, backscatter
- Medium risk - Serious incident, Resolve / reply within 2 days
  - attacks on Secure Shell (SSH) and Remote Desktop Protocol (RDP)
  - compromised machines, all kinds of botnets
  - potential threats to infrastructure - OpenNTP, HeartBleed, Network Address Translation- Port Mapping Protocol (NAT-PMP)
- High risk - Very serious event, Resolve / answer as soon as possible
  - DoS / DDoS (verified reports, not all received automated messages)

- incorrect configuration enables powerful attacks (DoS Scan Monitor Network Time Protocol (NTP), Simple Network Management Protocol (SNMP), Simple Service Discovery Protcol (SSDP))
- compromised network infrastructure.

Reports will be disseminated in a variety of formats and channels including email, HTTPS, RSS – the exact set to be determined. Report dissemination will be subject to compliance checking as described elsewhere in this document.

3.3.5 TI Sharing Compliance Monitor and Enforcer

The purpose of the compliance module is to check that TI shared complies with necessary sharing rules. More specifically, each rule (to abide by) can be described as a ***policy***, and a collection of policies can make up a *profile* to serve a particular compliance purpose, for instance to be *GDPR compliant* when they use PROTECTIVE, an NREN would need to apply a *GDPR profile.* The profile in that case is a collection of policies describing GDPR rules used to censor data leaving the NREN. These censoring rules will include anonymization, pseudonymization and aggregation of data in the TI, and in some cases may involve also dropping the TI altogether. The profile will be applied to the compliance module and check each TI event for violations.

It should be noted that any GDPR profile created will not be officially recognised or approved by the EU, it will simply be a demonstration of an organisation's effort to be GDPR compliant. A *default* GDPR profile will exist with each instance of PROTECTIVE. However, it should be noted that the GDPR profile should be reviewed in-depth by each NREN before using the PROTECTIVE tool within their organisation. This needs to be done to ensure GDPR compliance is still satisfied for their particular organisation. The addition of a default GDPR profile serves to minimise time spent on configuration, but also allow for the fundamental set of rules to be continually peer-reviewed by its users, and over time improve to ensure that one organisation's past mistakes can be corrected for and shared across various PROTECTIVE communities. A NDA profile as well as an organisation-specific profile can also be created to enforce policies that relate to NREN stakeholders or the NREN itself. This would be implemented in a similar manner to a GDPR profile.

Each policy is described by using a basic grammar consisting of a **name** (of the policy), the **rule** it abides by (typically a conditional check) and an **action** that makes up what the enforcer should do if the rule conditions for breaking the policy are met. This approach is akin to creating misuse-detection signatures in IDSs such as Snort[25]. Profiles allow policies to be interpreted by the module. These policies are stored in a persistent storage, as shown in Figure 18.

Our implementation of the compliance module will be an extension of a Mentat internal tool. Currently, CESNET's Inspector monitors values in incoming events and checks these against a list of rules. In the first instance of the compliance module, we assume that PROTECTIVE will:

- **Monitor and enforce policies on outgoing TI** (i.e. TI that is leaving an NREN, as opposed to incoming, which is what the Inspector currently does). This means that there will be no monitoring or enforcement on received TI. (In other words, informally: "I am only checking that I am not violating any regulations or legislation violations, but I am not checking my neighbour's violations"). This is done in the effort of simplicity – keeping track of one's own profiles is not as technically challenging as keeping track of many neighbours' profiles. In addition to requiring version control, it would be necessary to deal with distributed networks of profiles. Furthermore, there may be political, legislative, regulative or competitive reasons for why profiles should not be shared. We expect analysts to be responsible and report violations directly in case of any

[25] https://www.snort.org/

suspicion that a "neighbour's" TI violates any regulations or legislation.

- **Check (only) TI that uses PROTECTIVE channels as means of transportation**. This means that messages (TI) send by the analyst through an alternative means of communication not connected directly to PROTECTIVE, such as over an email ticketing system, will not be verified by the compliance checks. Therefore it would be impossible to determine automatically if that TI is or is not violating any regulations or legislation.
- **Need to investigate further on how data and information can be considered "sensitive" and in what contexts**. In order to understand when data and information (e.g. IP address) is considered private or sensitive, it is necessary for us to test several use case scenarios during the implementation of the compliance module and its various rules. Currently, we believe multiple Inspectors could be executed in tandem, or hierarchies of policies will need to be written to identify sensitivity of data and information.

Specifically, we consider two levels of censorship that can be outlined in the profiles:

- **Recipients** – who is allowed to see the data event (if any at all)?
- **Key-Value pairs** describing TI – how should keys and values be shown in the IDEA event (if at all)?

We are also taking inspiration from the IEP language. The IEP language can be used to develop the policies for the compliance module. The IEP provides an in-depth description of policy types as well as their enumerations, including the use of **Handling, Action, Sharing** and **Licensing** (HASL). Specifically, FIRST[26] writes: "*Handling policy statements define any obligations or controls on information received, to ensure the confidentiality of information that is shared. Action policy statements define the permitted actions or uses of the information received that can be carried out by a recipient. Sharing policy statements define any permitted redistribution of information that is received and any actions that need to be taken first. Licensing policy statements define any applicable agreements, licenses, or terms of use that governs the information being shared. For example, a reference to an existing partner sharing agreement or commercial license.*"

Current limitations of the IEP design prevent us from being fully IEP compliant. For instance, while we should be able to support **Sharing** as described by IEP, we are unlikely to support monitoring of other partners' activities – if we did, we believe the PROTECTIVE tool would turn into a policing tool that continually monitors what others do to the TI shared. We want to foster an environment not driven by data sharing paranoia, but one where responsible, automated data sharing rules are in place. These will protect privacy of data subjects, directly address ethical concerns, and will allow participants to raise concerns through the communities they belong to if automated aspects of data sharing were to fail.

We are likely to mainly use the latter two policy types: Sharing and Licensing. Sharing to describe restrictions and conditions for how TI can be shared, and license to be able to provide reference to which licences the policy has to abide by. This allows PROTECTIVE to specify sharing restrictions and conditions explicitly. IEPs come in a JSON format. These will be specified by an analyst, loaded in the persistent storage (see Figure 18) and used by the compliance module, akin to the Listing 326[26].

[26] https://www.first.org/iep/FIRST_IEP_framework_1_0.pdf

```
"FIRST-mailing-list-iep":{
        "id":"01bc4353-4829-4d55-8d52-0ab7e0790df9",
        "name":"FIRST.orgMailingListIEP",
        "version":1,
        "reference":"https://www.first.org/mailing-list-iep",
        "start-date":"2016-06-0910:09:00",
        "end-date":"2016-12-3110:09:00",
        "encrypt-in-transit":"MAY",
        "encrypt-at-rest":"MAY",
        "permitted-actions":"EXTERNALLYVISIBLEDIRECTACTIONS",
        "affected-party-notifications":"MAY",
        "sharing-level":"AMBER",
        "attribution":"MUSTNOT",
        "obfuscate-affected-parties":"MUST",
        "unmodified-resale":"MUSTNOT",
        "external-reference":"https://www.first.org/about/policies/bylaws"
}
```


Listing 3: Example of IEP

Key and values are akin to those found in the IDEA JSON schema. In the example from CESNET's IDEA webpage[27]: **`"IP4": ["192.168.0.2-192.168.0.5","192.168.0.10/25"]`** we see the key "IP4" indicates the "row" of IP4 values (i.e. remove or modify the key and all values have to relate to the new key or get removed). Whereas values are individual instances of IP4 (in this example). Modify or remove that one, the key will still stand, and so will all other values.

Specific actions the compliance manager aims to support include:

- **Drop** TI event from being sent;
- **Substitute** (values in TI) using:
    - **Anonymization**[28] – i.e. remove identifying particulars or details from values
    - **Pseudonymization**[20] – i.e. replacing values with artificial identifiers
    - **Aggregation**[20] - i.e. abstract out or otherwise summarise data to preserve privacy
    - **Correction** – i.e. assuring well-formedness
    - **Deletion** – i.e. removal of values
- **Report** TI event having been modified in some way (intended for documentation purposes).

As mentioned, we recognise that no compliance module will be able to identify all (true positive) violations correctly every time (we would expect some false negatives and false positives). Should sensitive or personal data be leaked, it is necessary to have specified appropriate high-level (human-level) PROTECTIVE partner protocols in place to address automation or human error (should the compliance module have false negative violations have been shared). Before using the PROTECTIVE tool, we will ask NRENs and SMEs to join in on an IEP that outlines best practices, common courtesy and protocols for use of the tool. The details of these human-level rules for using the tool are still being outlined, and are expected to be completed before the end of the project[29].

[27] https://idea.cesnet.cz/en/index
[28] Refers to a GDPR action
[29] For an in-depth description of the project's ethical and data protection plan, please review deliverable D2.4.

4 Conclusion

The deliverable described the PROTECTIVE TI sharing. We outlined how TI sharing is based on the state of the art, and have also identified areas in which we can innovate. We first highlighted ambiguities in the literature on TI, then presented basic building blocks to consider for our platform. Moreover, we presented a critical analysis of the state of the art in the area of information types, and models, methods and mechanisms for TI sharing. This knowledge (as well as some various corresponding legal aspects) is the basis for the specifications of PROTECTIVE's architecture. In more details, the document described both high-level picture of the architecture and also its individual components.

5 References

ACTRA. (2012). *Arizona Cyber Threat Response Alliance* . Retrieved 2017, from http://azinfragard.org/?page_id=8

Ahrend, J. M., Jirotka, M., & Jones, K. (2016). On the collaborative practices of cyber threat intelligence analysts to develop and utilize tacit Threat and Defence Knowledge. Cyber Situational Awareness, Data Analytics And Assessment (CyberSA).

Aviram, A., & Tor, A. (2004). *Overcoming Impediments to Information Sharing.* Harvard Law and Economic Discussion Paper.

Barnum, S. (2012). *Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX).* MITRE. Retrieved from Barnum, S. (2012). Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIX™). MITRE Corporation, 11.

Boulding, K. (1955). Notes on the Information Concept. *Exploration*, 21–32.

Boyens, J., Paulsen, C., Moorthy, R., Bartol, N., & Shankles, S. A. (2014). *Supply chain risk management practices for federal information systems and organizations.* NIST.

Bromiley, M. (2016). *Threat Intelligence: What It Is, and How to Use It Effectively SANS.* SANS.

Burger, E. W., Goodman, M. D., Kampanakis, P., & Zhu, K. A. (2014). Taxonomy model for cyber threat intelligence information exchange technologies. ACM Workshop on Information Sharing & Collaborative Security.

Burne, C., Norman, T. J., & Sycara, K. (2013). Stereotypical trust and bias in dynamic multiagent systems. *ACM Transactions on Intelligent Systems and Technology (TIST), 4*(2), 22.

Calder, A., & Watkins, S. (2008). IT governance: A manager's guide to data security and ISO 27001/ISO 27002. Kogan Page Ltd.

CERT-UK. (2015). Integrating Threat Intelligence Defining an Intelligence Driven Cyber Security Strategy. CERT-UK, CPNI.

Chismon, D., & Ruks, M. (2015). *Threat Intelligence: Collecting, Analysing, Evaluating.* MWR InfoSecurity.

Connolly, J., Davidson, M., & Schmidt, C. (2014). *The trusted automated exchange of indicator information (TAXII).* Mitre.

Dalziel, H. (2014). *How to define and build an effective cyber threat intelligence capability.* Syngress.

Dandurand, L., & Serrano, O. S. (2013). Towards improved cyber security information sharing. International Conference on Cyber Conflict (CyCon).

Daniel, F., Casati, F., D'Andrea, V., Mulo, E., Zdun, U., Dustdar, S., . . . al., e. (2009). *Business compliance governance in service-oriented architectures.* IEEE International Conference on Advanced Information Networking and Applications.

Danyliw, R., Meijer, J., & Demchenko, Y. (2007). *The incident object description exchange format*. Retrieved 2017, from https://tools.ietf.org/html/rfc5070

Debar, H., Curry, D., & Feinstein, B. (2007). *Intrusion Detection Message Exchange Format (IDMEF)*. Retrieved 2017, from https://www.ietf.org/rfc/rfc4765.txt

England, B. o. (2016). *CBEST Intelligence-Led Testing Understanding Cyber Threat Intelligence Operations.* Bank of England.

ENISA. (2013). *Detect, SHARE, Protect - Solutions for Improving Threat Data Exchange among CERTs.* ENISA https://www.enisa.europa.eu/publications/detect-share-protect-solutions-for-improving-threat-data-exchange-among-certs.

ENISA. (2014). *Standards and tools for exchange and processing of actionable information*. Retrieved 2017, from https://www.enisa.europa.eu/publications/standards-and-tools-for-exchange-and-processing-of-actionable-information/

ENISA. (2016). *A good practice guide of using taxonomies in incident prevention and detection.* ENISA.

ENISAa. (2014). *Actionable Information for Security Incident Response*. Retrieved from https://www.enisa.europa.eu/publications/actionable-information-for-security/

ENISAa. (2016). *ENISA Threat Taxonomy*. Retrieved 2017, from https://www.enisa.europa.eu/topics/threat-risk-management/threats-and-trends/enisa-threat-landscape/etl2015/enisa-threat-taxonomy-a-tool-for-structuring-threat-information

ENISAb. (2016). *NCSS Good Practice Guide Designing and Implementing National Cyber Security Strategies*. Retrieved 2017, from https://www.enisa.europa.eu/publications/ncss-good-practice-guide

ENISAb. (2016). *NCSS Good Practice Guide Designing and Implementing National Cyber Security Strategies*. Retrieved 2017, from https://www.enisa.europa.eu/publications/ncss-good-practice-guide

FireEye. (2017). *FireEye Threat Intelligence Webpage https://www.fireeye.com/products/cyber-threat-intelligence.html.* FireEye.

Fransen, F., Smulders, A., & Kerkdijk, R. (2015). *Cyber security information exchange to gain insight into the effects of cyber threats and incidents.* Elektrotechnik und Informationstechnik.

Frické, M. (2009). The knowledge pyramid: a critique of the DIKW hierarchy. *Journal of information science*, 131-142.

Friedman, J., & Bouchard, M. (2015). *Definitive guide to cyber threat intelligence.* CyberEdge Press.

Garrido-Pelaz, R., González-Manzano, L., & Pastrana, S. (2016). Shall we collaborate?: A model to analyse the benefits of information sharing. ACM Workshop on Information Sharing and Collaborative Security.

Goodwin, C., Nicholas, J. P., Bryant, J., Ciglic, K., Kleiner, A., Kutterer, C., & Storch, T. (2015). *A framework for cybersecurity information sharing and risk reduction.* Microsoft.

Haass, J., Ahn, G.-J., & Grimmelmann, F. (2015). ACTRA-A Case Study for Threat Information Sharing. ACM Workshop on information sharing and collaborative security.

Habib, S. M., Ries, S., Hauke, S., & Mühlhäuser, M. (2012). Fusion of Opinions under Uncertainty and Conflict -- Application to Trust Assessment for Cloud Marketplaces. *IEEE 11th International*

Conference on Trust, Security and Privacy in Computing and Communications (TrustCom), pp. 109-118.

Habib, S. M., Volk, F., Hauke, S., & Mühlhäuser, M. (2015). Computational Trust Methods for Security Quantification in the Cloud Security Ecosystem. In R. Ko, & K.-K. R. Choo (Eds.), *The Cloud Security Ecosystem: Technical, Legal, Business and Management Issues* (pp. 463-493). Syngress/Elsevier.

Harkins, M. (2016). *Managing risk and information security.* Apress.

Howard, J. D., & Longstaff, T. A. (1998). *A common language for computer security incidents.* Sandia National Laboratories.

Jang, J.-w., Kang, H., Woo, J., Mohaisen, A., & Kim, H. K. (2015). Andro-AutoPsy: Anti-malware system based on similarity matching of malware and malware creator-centric information. *Digital Investigation, 14*, 17-35.

Johnson, C., Badger, L., Waltermire, D., Snyder, J., & Skorupka, C. (2016). *Guide to Cyber Threat Information Sharing.* National Institute of Standards and Technology (NIST).

Kacha, P. (2013). IDEA: Designing the Data Model for Security Event Exchange. Computers: Recent Advances in Computer Science.

Kacha, P. (2014). IDEA: Security Event Taxonomy Mapping. Circuits, Systems, Communications and Computers: Advances in Information Science and Applications.

Kacha, P. (2015). IDEA: Classification of security events, their participants and detection probes. *WSEAS Transactions on Computers*.

Kacha, P., M. Kostenec, M., & Kropacova, A. (2016). Warden 3: Internet Threat Sharing Platform. *International Journal of Computers*.

Kaijankoski, E. A. (2015). *Cybersecurity Information Sharing Between Public-Private Sector Agencies.* Calhoun.

Kijewski, P., & Pawliński, P. (2014). Proactive Detection and Automated Exchange of Network Security Incidents. Abgerufen am.

Lewis, R., Louviens, P., Abbott, P., Clewley, N., & Jones, K. (2014). Cybersecurity Information Sharing: A Framework for Sustainable Information Security Management in UK SME Supply Chains. European Conference on Information Systems (ECIS).

Lievesley, D. (2006). Data information knowledge chain. *Health Informatics Now*.

Liu, Y., Muller, S., & Xu, K. (2007). A static compliance-checking framework for business process models. *IBM Systems Journal, 46*, 335-361.

MACCSA. (2013). *Information Sharing Framework*. Retrieved March 14, 2017, from https://www.terena.org/mail-archives/refeds/pdfjJz1CRtYC4.pdf

Mauro, F., & Stella, D. (2016). Brief Overview of the Legal Instruments and Restrictions for Sharing Data While Complying with the EU Data Protection Law. International Conference on Web Engineering.

McMillan, R. (2013). *Definition: Threat Intelligence*. Retrieved March 2, 2017, from https://www.gartner.com/doc/2487216

Mohaisen, A., Al-Ibrahim, O., Kamhoua, C., Kwiat, K., & Njilla, L. (2017). *Rethinking information sharing for actionable threat intelligence.* Cornell University Library.

Moriarty, K. (2012). *Real-time inter-network defense.*

Movius, L., & Krup, N. (2009). US and EU privacy policy: comparison of regulatory approaches. *International Journal of Communication*.

NCIAgency. (2017). *Malware Information Sharing Platform*. Retrieved 2017, from https://www.ncia.nato.int/Documents/Agency%20publications/Malware%20Information%20Sharing%20Platform%20(MISP).pdf

Obrst, L., Chase, P., & Markeloff, R. (2012). Developing an Ontology of the Cyber Security Domain. STIDS.

O'Leary, D. E., Bonorris, S., Klosgen, W. K., Lee, H. Y., & Ziarko, W. (1995). Some privacy issues in knowledge discovery: the OECD personal privacy guidelines. *IEEE Expert, 10*(2), 48 - 59.

Paxson, V. (1999). Bro: A System for Detecting Network Intruders in Real-Time. *Computer Networks*, 2435-2463.

Peretti, B. (2014). *Hearing on Cyber Security: Prepared Testimony.* U.S. Senate Committee on Banking, Housing, and Urban Affairs.

Ries, S., Habib, S. M., Mühlhäuser, M., & Varadharajan, V. (2011). CertainLogic: A Logic for Modeling Trust and Uncertainty (Short paper). *4th International Conference on Trust and Trustworthy Computing*, pp. 254-261.

Roesch, M. (1999). Snort: Lightweight intrusion detection for networks. *Lisa, 99*(1), 229-238.

Rowley, J. (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information and Communication Science*, 163–180.

Rowley, J., & Hartley, R. (2006). *Organizing Knowledge: An Introduction to Managing Access to Information.* Ashgate.

Sarbanes, P. (2002). *Sarbanes-oxley act of 2002.* The Public Company Accounting Reform and Investor Protection Act. Washington DC: US Congress.

Sauerwein, C., Sillaber, C., Mussmann, A., & Breu, R. (2017). Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives. Wirtschaftsinformatik.

Sedenberg, E. M., & Mulligan, D. K. (2015). *Public Health as a Model for Cybersecurity Information Sharing.* Berkeley .

Serrano, O., Dandurand, L., & Brown, S. (2014). On the design of a cyber security data sharing system. ACM Workshop on Information Sharing & Collaborative Security.

Sillaber, C., Sauerwein, C., Mussmann, A., & Breu, R. (2016). Data Quality Challenges and Future Research Directions in Threat Intelligence Sharing Practice. ACM Workshop on Information Sharing and Collaborative Security.

Soto-Mendoza, V., Serrano-Alvarado, P., Desmontils, E., & Garcia-Macias, J. (2015). Policies composition based on data usage context. International Workshop on Consuming Linked Data (COLD2015) at ISWC.

TeleManagement Forum. (2013). *Sharing Threat Intelligence to Mitigate Cyber Attacks*. Retrieved 2017, from https://www.edge-technologies.com/system/files/documents/SharingThreatIntelligence_ArchitectureV0.8final.pdf

Tripwire. (2014). *Tripwire Webpage https://www.tripwire.com/state-of-security/security-data-protection/cyber-threat-intelligence/.* Tripwire.

Vasek, M., Weeden, M., & Moore, T. (2016). Measuring the Impact of Sharing Abuse Data with Web Hosting Providers. ACM Workshop on Information Sharing and Collaborative Security.

Vasilomanolakis, E., Karuppayah, S., Fischer, M., Mühlhäuser, M., Plasoianu, M., Pandikow, L., & Pfeiffer, W. (2013). This network is infected: HosTaGe-a low-interaction honeypot for mobile devices. *ACM workshop on Security and privacy in smartphones & mobile devices*, pp. 43-48.

Vasilomanolakis, E., Karuppayah, S., Kikiras, P., & Mühlhäuser, M. (2015). A honeypot-driven cyber incident monitor: lessons learned and steps ahead. *International Conference on Security of Information and Networks*, pp. 158-164.

Wagner, C., Dulaunoy, A., Wagener, G., & Iklody, A. (2016). MISP-The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. WISCS.

Webroot. (2014). *Threat Intelligenve: What is it, and how can it protect you from today's advanced cyber-attacks?* Webroot.

Willis, B. (2012). *Sharing Cyber-Threat Information: An Outcomes-based Approach.* Intel Corporation.

Yann, L., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature* .

Zeleny, M. (2005). Human Systems Management: Integrating Knowledge, Management and Systems. *World Scientific*, 15–16.

Zhao, W., & White, G. (2012). *A collaborative information sharing framework for community cyber security.* Technologies for Homeland Security (HST).

Zhao, W., & White, G. (2014). Designing a formal model facilitating collaborative information sharing for community cyber security. System Sciences (HICSS).

Zins, C. (2007). Conceptual Approaches for Defining Data, Information, and Knowledge. *Journal of the American Society for Information Science and Technology, 58*(4), 479–493.