SPARTA

D4.1

Cybersecurity threat intelligence common data model

Project number	830892
Project acronym	SPARTA
Project title	Strategic programs for advanced research and technology in Europe
Start date of the project	1st February, 2019
Duration	36 months
Programme	H2020-SU-ICT-2018-2020

Deliverable type	Demonstrator
Deliverable reference number	SU-ICT-03-830892 / D4.1 / V1.0
Work package contributing to the deliverable	WP4
Due date	July 2020 – M18
Actual submission date	30th July, 2020

Responsible organisation	CESNET
Editor	Martin Žádník
Dissemination level	PU
Revision	V1.0

Abstract	There are many standards for sharing of operational cyber threat intelligence (CTI). The goal of this deliverable is to select a suitable candidate solution from the existing data models and provide its extensions and unification so it can become practical and unified data model.
Keywords	Data model, sharing, threat intelligence

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 830892.

Editor

Martin Žádník (CESNET)

Contributors (ordered according to beneficiary numbers)

Augustin Lemesle (CEA)

Giorgos Papavassiliou (KEMEA)

Juan Caubet (EUT)

Antonio Pagano, Mauro Gil Cabeza, Rumen Daton Madenou (IND)

Romain Ferrari, Olivier Bettan (TCS)

Paolo Mori, Oleksii Osliak (CNR)

Claudio Porretti (LEO)

Šarūnas Grigaliūnas (KTU)

Evaldas Bružė, Edmundas Piesarskas (L3CE)

Jocelyn Aubert (LIST)

Bertrand Lathoud, Tun Hirt (SMILE)

Pawel Pawlinski (NASK)

Filipe Apolinario (INOV)

Reviewers (ordered according to beneficiary numbers)

Mohammad Norouzian (TUM)

Alessio Merlo (CINI)

Executive Summary

This deliverable documents our activities, findings and deployment while investigating a common data model for sharing of operational cyber threat intelligence (CTI). First, we collected requirements from the partners involved in T-SHARK Programme of SPARTA. These requirements arise from the development activities of comprehensive threat analysis, sharing and visual analytics. The common denominator of these activities is a complex use case of election interference with its specific sub-cases. In parallel, we collected related research, existing frameworks and tools to see our topic from various perspectives. Subsequently, we analysed strong and weak points as well as opportunities and threats of the existing works. We identified MISP and STIX as promising candidates.

The results of our analysis serve as input for an expert workshop which makes a strategic decision to develop the data model using concepts available in MISP. The results also show that rather than the common data model, it is essential to introduce a methodology to create data models for specific use cases so that the data models are more consistent and exhibit better machine readability. Another result of the expert workshop is a vision of ontology that would enable mapping between the individual data models. We leave the latter result as our future work, then we introduce a methodology to prepare unified data models in MISP and we demonstrate our methodology on several selected sub-cases. Furthermore, we show its applicability in a production environment by deploying SPARTA MISP instance and extending it with the newly proposed data models. The demonstration is prepared using a fictional story of election interference and how the interference might reflect into CTI shared via MISP. We conclude the deliverable with future plans which arise from the activities related to the activities around the deliverable as well as from the activities within T-SHARK Programme.

Table of Content

Chapter 1 Introduction 1
Chapter 2 CTI related work 3
2.1 Security Vocabularies 3
2.2 CTI formats 4
2.3 CTI sharing and analysis platforms 6
2.4 Actionable CTI 8
2.5 Existing frameworks 8
2.5.1 NIST Cybersecurity Framework 8
2.5.2 NATO AC/35-D/1017 10
2.5.3 EBIOS 10
2.5.4 MITRE att&ck 10
2.5.5 ISO 11
2.5.6 ISA 13
2.5.7 EPRI 13
2.5.8 IEEE 13
2.5.9 ITU 13
Chapter 3 Use cases and requirements 14
3.1 Methodology 14
3.2 Use cases 14
3.3 Summary of requirements 17
Chapter 4 Strategic analysis 19
4.1 SWOT analysis 19
4.1.1 Background 19
4.1.2 SWOT of CTI common data model 20
4.1.3 SWOT conclusions and convergence 24
4.2 Expert workshop 25
4.2.1 Methodology 25
4.2.2 Summary of the work session 26
4.2.3 Identified key requirements for the data model 27
4.2.4 Next Steps 28
Chapter 5 Unification of data models 29
5.1 Methodology 29
5.2 Data models 32

5.2.1 DDoS backscatter 32
5.2.2 Twitter 36
5.2.3 Election interference 39
5.2.4 BP-IDS 41
5.2.5 Malware 44
Chapter 6 Demonstration 48
6.1 Deployment 48
6.2 Demonstration of election interference 49
6.2.1 The fictional story of election interference 49
6.2.2 Reflection in MISP 50
Chapter 7 Plans and roadmap 68
Chapter 8 Summary and Conclusion 71
Chapter 9 Bibliography 72
Chapter 10 Annex – A: Requirements 74
Chapter 11 Annex – B: Raw combined SWOT Matrix 90

List of Figures

Figure 2.1: NIST Cybersecurity Framework 9
Figure 2.2: The EBIOS risk management rocess 10
Figure 2.3: Part of MITRE Att&ck matrix 11
Figure 2.4: Risk assessment process according to ISO/IEC 27005 12
Figure 4.1: Graphical representation of the SWOT analysis results 20
Figure 4.2: Templates for distributed SWOT analysis 21
Figure 6.1: DDoS event step 1 51
Figure 6.2: DDoS event step 2 51
Figure 6.3: DDoS event step 3 52
Figure 6.4: DDoS event step 4 52
Figure 6.5: Full DDoS event 53
Figure 6.6: DDoS event proposal step 1 54
Figure 6.7: DDoS event proposal step 2 54
Figure 6.8: Spam event 55
Figure 6.9: Twitter event 56
Figure 6.10: Spam event related with the Twitter event 57
Figure 6.11: Election interference event 57
Figure 6.12: Editing DDoS event to extend election interference event 59
Figure 6.13: Additional events linked with election interference event 59
Figure 6.14: Extended view of event with extending events 60
Figure 6.15: List of events extending election interference event 61
Figure 6.16: BP-IDS alert event 62
Figure 6.17: BP-IDS event enriched 63
Figure 6.18: BP-IDS alert updated 63
Figure 6.19: Malware event of the first sample 64
Figure 6.20: Malware event enriched 65
Figure 6.21: Malware event of the second sample 66
Figure 6.22: List of all events 67

List of Tables

Table 2.1: STIX Objects 4
Table 2.2: MAEC Objects 5
Table 2.3: Sharing platforms 7
Table 4.1: Research and industrial entities expert in Cyber Threat Intelligence 22
Table 10.1: Requirements 74
Table 11.1: Combined SWOT 90

Chapter 1 Introduction

Today, cyberspace is widely recognised as the fifth operational domain, joining the traditional fields of land, sea, air, and space. This acknowledges that the battlefield has expanded to include attacks over the internet and via the electromagnetic spectrum, which can destroy infrastructure, manipulate or interfere with critical national databases, and even cause physical damage.

Cyber-attacks are becoming more complex over time and affect multiple domains. An attack that starts from the web can become a physical attack, and vice versa. For example, after a physical intrusion inside a building, the attacker hides a network device inside the building so that it can subsequently attack the building remotely.

The most common attacks such as defacement, DDoS, or the exploitation of weaknesses of computer systems are favoured by the fact that much of the information is exposed online or sold on black markets as exploits or easy-to-use tools. Targeted attacks by criminal or state-sponsored organisations about which it is difficult to obtain information are also increasing. The majority of these attacks start from Social Engineering activities, hitting the profiles of public figures or employees and then hitting the organisations as the final victim themselves.

The options of an attacker and its defender in cyberspace are asymmetric. The attacker chooses the time, the space as well as the mean. Moreover, the cost of an attack is low in comparison with the cost of adequate defence measures. A way to reduce this imbalance is cooperation among defenders such as exchange of cyber threat intelligence.

Cyber threat intelligence plays an important role in everyday life of cybersecurity practitioners to receive new information about threats, vulnerabilities, attacks, indicators of compromise as well as it supports tactical decisions, improves available defence mechanisms and introduces new strategies for mitigating or even preventing the assets from cyberattacks. Unfortunately, the biggest challenge is to make use of information, i.e., how to comprehend the information and implement its remedy, since CTI is not simply information, it is information that has been analysed and is actionable [1]. Besides, automating the process of CTI sharing, even the consumption itself has raised new challenges for researchers and practitioners. Current threat intelligence platforms provide limited mechanisms for automation [2].

There are quite a few various data models available as of today, some of them complex, some of them single purpose, some of them accepted while some of them obsolete. Indeed, there is no data model that is widely accepted across the world, nor in EU nor in individual countries.

One of the SPARTA work package 4 (aka T-SHARK Programme) goals is to lay the groundwork for the common cybersecurity threat intelligence data model to improve capabilities of the team to produce and consume threat intelligence and to share the intelligence within and among heterogeneous communities.

Therefore, we narrow down our investigation of the common data model to sharing of operational CTI in this deliverable having in mind that

- Iterative expansion of a common but dedicated model will more likely be adopted than a complex model capturing all aspects of CTI.
- Community adoption is crucial to ensure the development of the model beyond SPARTA.
- Sharing of CTI is part of T-SHARK (Task 4.4 of SPARTA); hence we can leverage the competence of involved partners.

We intend to come up with a methodology for creating data models that will be capable of covering operational CTI data from several intelligence domains: technical, social, information, human and physical domain, as well as it will express relationships between these data. The structured information in a standardised format will improve machine readability and sharing.

This deliverable captures our steps in investigating the CTI common data model. In the beginning, we narrow down our investigation from the general CTI to the data model relevant for sharing of operational CTI. Further, Survey of existing vocabularies, frameworks, tools and models is provided in Chapter 2. Subsequently, the deliverable documents requirements drawn from our versatile use cases in Chapter 3. Chapter 4 captures our strategic choice over the existing tools and models. We define a methodology to achieve the common data model in Chapter 5, together with its application on the subcases. We demonstrate practical pilot implementation in Chapter 6. Chapter 7 outlines the roadmap of further development and long-term plans.

Chapter 2 CTI related work

Many organisations produce, collect and share information related to potential and known cyberattacks. According to the National Institute of Standards and Technology (NIST), cyber threat information is any information related to threats that might help organisations in protecting themselves against cyberattacks or in detecting the activities of adversaries. While Cyber Threat Intelligence (CTI[1]), is what threat information becomes after its processing and analysis. Another definition given by Gartner considers CTI as evidence-based knowledge that includes context, mechanisms, indicators, implications and actionable advice about an existing or emerging menace or hazard to IT or information assets. Many organisations (e.g., NIST[2], MITRE[3]) have formulated enumerations of types of malware, vulnerabilities, and exploitations. Particularly, MITRE maintains three dictionaries, namely Common Vulnerabilities and Exposures (CVE[4]), Common Platform Enumeration (CPE[5]) and Common Weakness Enumeration (CWE[6]).

2.1 Security Vocabularies

The CVE was launched in the late ’90s to address the lack of a list of standard identifiers for known vulnerabilities since the majority of cybersecurity tools relied on their databases with their names for vulnerabilities. Furthermore, another security gap was related to the lack of a standardised basis for evaluation among tools since each tool vendor used different metrics to declare the number of vulnerabilities or exposures they detected. Differently to the CVE that mainly deals with specific instances within a product or system, the CWE defines a list of standard software and hardware weakness types. According to the official definition given by MITRE, the CWE defines errors (e.g., vulnerabilities, bugs) in the implementation, code, or architecture in software or hardware. Thus, system, network, or hardware will be vulnerable to cyberattack if such errors will remain. In its turn, the CPE provides an XML-based dictionary that follows the structured scheme for naming information technology systems, software, and software packages, thus providing a common representation of a specific software product including product name, vendor, and the product version.

Information provided by the dictionaries mentioned above is important; however, none of these dictionaries describes the ways of how adversaries exploit weaknesses of the system or software. Therefore, in 2007, MITRE Corporation released a Common Attack Pattern Enumeration and Classification (CAPEC[7]) dictionary. The CAPEC dictionary defines a list of descriptions, known as Attack Patterns of the common attributes and approaches used by adversaries to exploit known weaknesses. Each attack pattern provides knowledge about how specific phases and components of an attack are designed and executed. Furthermore, an attack pattern may contain guidelines for mitigating the effects of an attack.

[1] We use the abbreviation CTI to refer to cyber threat intelligence and not to cyber threat information throughout this document.
[2] https://www.nist.gov/
[3] https://www.mitre.org/
[4] https://cve.mitre.org/
[5] https://cpe.mitre.org/
[6] https://cwe.mitre.org/
[7] https://cape.mitre.org/

2.2 CTI formats

Several approaches for CTI sharing were defined in the last years. The OpenIOC[8] is an extensible XML scheme for the description of technical characteristics that identify known threats, the methodology used by the threat agent, or other evidence of compromise. However, OpenIOC suffers from the lack of supporting techniques, tactics and procedures (TTPs) description and it has a limited commercial adoption comparing to other standards. Incident Object Description Exchange Format (IODEF) defines an XML data representation that provides a framework for sharing information commonly exchanged by Computer Security Incident Response Teams (CSIRTs) about computer security incidents. IODEF is an open standard through the Internet Engineering Task Force (IETF). However, it requires other formats for describing TTPs and campaigns. Furthermore, IODEF was designed to share incident data instead of IoCs. Collective Intelligence Framework (CIF[9]) is the open-source platform used to store and to share CTI. It utilises the IODEF data format for sharing and storing threat-related information. The most common types of threat intelligence stored in CIF are IP addresses, domains, and URLs related to malicious activities. This framework covers various data-observations from any source and creates a series of observations. Meanwhile, CIF does not provide a description of TTPs and threat actor data.

Structured Threat Information Expression (STIX[10]) language was presented together with a transport mechanism for secure CTI sharing known as Trusted Automated Exchange of Intelligence Information (TAXII[11]). While TAXII is a protocol used to exchange CTI over HTTPS, the main goal of the STIX standard is allowing organisations to exchange CTI in a consistent and machine-readable format, thus allowing security communities to respond to cyberattacks timely and more effectively. Currently, there are two major versions of the language existing online, i.e., STIX1.x and STIX2.x. The STIX2.x follows the JSON format, while STIX1.x was defined using XML. Moreover, differently to the STIX1.x, the STIX2.x defines 18 top-level objects called STIX Domain Objects (SDOs) and each of them corresponds to a specific concept of the CTI. Furthermore, the generic TTPs and Exploit Target types were split into separate top-level objects Attack Pattern, Malware, Tool and Vulnerability. Additionally to SDOs, the STIX2.x standard defines top-level STIX Relationship Objects (SROs) that define relations between SDOs by linking them by named relationship types. Table 2.1 reports STIX objects with their brief description.

Table 2.1: STIX Objects

Object name	Description
Observed Data	Conveys information about cybersecurity related entities such as files, systems, and networks.
Attack Pattern	Belongs to TTPs that describe ways that adversaries attempt to compromise targets.
Campaign	A grouping of adversarial behaviours that describes a set of malicious activities or attacks.
Indicator	Contains a pattern for detecting a suspicious or malicious activity.
Malware	Belongs to TTPs and represents malicious code.
Malware Analysis	The metadata and results of a particular static or dynamic analysis performed on malware.

[8] http://www.openioc.org/
[9] http://csirtgadgets.org/
[10] https://oasis-open.github.io/cti-documentation/stix/intro.html
[11] https://oasis-open.github.io/cti-documentation/

Object name	Description
Tool	Legitimate software that can be used by adversaries to initiate and perform attacks.
Vulnerability	A mistake in software that can be used by an adversary to gain access to a system or network.
Course of Action	A recommendation from a producer of intelligence to a consumer for mitigating and/or preventing an attack.
Identity	Individuals, organisations, or groups as well as classes of individuals, organisations, systems or groups.
Threat Agent	Individuals, groups, or organisations believed to be operating with malicious intent.
Infrastructure	Belongs to TTPs and describes a system, software service and any associated physical or virtual resources used by adversaries.
Intrusion set	Describes a set of adversarial behaviours and resources with common properties used by a single organisation.
Opinion	Describes a textual assessment of the information correctness in a STIX Object produced by a different entity.
Location	Represents a geographic location.
Report	A collection of CTI focused on one or more topics.
Note	Contains informative text to provide further context or additional analysis not contained in the STIX Objects.
Grouping	Asserts that the referenced STIX Objects have a shared context.
Relationship	Link together two SDOs or SCOs to describe their relation.
Sighting	Denotes the belief that something in CTI was seen.

Additionally, to the STIX standard that describes general concepts of the CTI, at the beginning of 2011, MITRE introduced the Malware Attribute Enumeration and Characterisation (MAEC) language for sharing and encoding of high-fidelity information about malware. The MAEC language aims to eliminate the inaccuracy and uncertainty of malware descriptions. MAEC aims to transform malware research and response by improving communication and reducing potential malware analysis duplication. Similarly, to the STIX standard, the MAEC defines several top-level objects: Behaviours, Malware Actions, Malware Families, Malware Instances, and Collections. The data model of MAEC is represented as a connected graph of nodes and edges, where MAEC top-level objects define the nodes and MAEC relationships define the edges as links between MAEC objects are reported in Table 2.2 with a short description.

Table 2.2: MAEC Objects

Object name	Description
Behaviours	Corresponds to a specific purpose behind a particular snippet of code, as executed by a malware instance.

Object name	Description
Malware Actions	Represents an abstraction on a system-level API call called by the malware instance during its execution.
Malware Families	Defines a set of malware instances that are related by common authorship or lineage.
Malware Instances	A single member of a Malware Family packaged as a binary.
Collections	Captures a set of MAEC entities or STIX Cyber Observables that are related or associated.

Objects both in STIX2.x and MAEC languages are defined through the set of corresponding attributes called properties that convey specific information.

In 2013, MITRE began developing the ATT&CK framework to accumulate knowledge about known cyberattacks and particularly descriptions of tactics and techniques used by adversaries. Furthermore, the ATT&CK framework provides a textual description of actions to be taken for resolving security issues, e.g. vulnerability patching, firewall configuration, applying encryption, etc. The framework uses STIX2.x standard for describing CTI thus can be shared by using tools and standards that support CTI sharing.

However, a recent study [3] indicates that STIX and TAXII had attracted interest in 18 countries, but their adoption presented some barriers. The specific barriers are initial setup and learning curve; organisational compatibility and maturity; understanding of cyber threat vocabulary; and a lack of conformity in notating data. On the contrary, specific benefits of adoption were enhanced sharing of structured relationship data; data restriction enabling; structured documentation mark-up; and improved interoperability.

Therefore, although many standards and data formats for a comprehensive description of CTI were proposed, the need for extending those approaches still exists and is recognised in [4]. Some works on extending the STIX standard were proposed. Thus, in [5] authors presented an extension for the STIX language that allows describing complex patterns. By using the proposed extension, security specialists can tag attributes of an object and use them for describing precise relations between different objects. However, the proposed extension applies only for the XML-based version of the STIX language, while the latest STIX version defines multiple relations between cyber-observable objects and other STIX domain objects. Another extension of the STIX standard was proposed in [6]. It provides a representation of the Data-Sharing Agreement describing both data controller and processor information, together with actions to be enforced before sharing CTI reports, thus satisfying GDPR constrains. The proposed extension was validated by enforcing the anonymisation mechanism on spam-emails represented as the proposed custom STIX object. Finally in [7], the authors proposed an extension to the STIX standard that allows describing sticky policies as a package of multiple custom STIX objects. The work presents a custom STIX bundle with predefined attributes for describing the validity period of the policy. Additionally, the authors proposed two specific custom objects for describing conditions for restricting usage of CTI reports and requirements to be enforced before sharing those CTI reports. Thus, the proposed extension can be shared with other STIX objects as a single bundle or as a separate document. The proposed extension was validated with the designed tool that allows writing sticky policies and enforcing specified anonymisation action in an automated manner.

2.3 CTI sharing and analysis platforms

A general CTI sharing platform typically provides the CTI creation, collection, exchange, and analytical capabilities within one or multiple communities. Furthermore, it may provide automated dissemination or enforcement of actionable CTI concepts such as courses of actions in order to

prevent or detect cyberattacks. Moreover, organisations can use multiple sharing platforms to exchange CTI within different levels, for example, between communities, organisations. However, existing platforms use different terms for describing the same concept, such as a CTI record, an Event in MISP, Pulse in OTX, an Activity in IBM X-Farce Exchange. Table 2.3 reports some of the well-known sharing platforms.

Table 2.3: Sharing platforms

Product	Vendor	Description
Malware Information Sharing Platform	Open Source	Free and open source community sourced CTI sharing platform
ThreatConnect	ThreatConnect	General CTI platform with community sharing capabilities
Cyber Threat Exchange	NC4	CTI sharing platform used by the FS-ISAC
Blueliv Threat Intelligence Platform	Blueliv	General CTI platform with community sharing capabilities

As the CTI management field is still in development, the functionality and features of existing CTI sharing platforms may change in the near future. Different CTI sharing platforms follow their philosophy, terminology, features, and focus on specific cyber threat data.

However, in recent years, MISP became de-facto a standard approach for collecting and sharing of CTI. MISP itself is a community sharing platform that depends on the content which is generated by communities. The platform supports CTI sharing via the web interface or Python library and a hub-spoke sharing community can be set up. Additionally, MISP has a protocol used for synchronisation between different MISP instances. The platform supports several synchronisation mechanisms including pull, push, and cherry-picking. While the pull mechanism allows one MISP to discover events of another MISP instance based on predefined distribution rights, the push mechanism allows sending single or multiple records to a remote instance. Finally, the cherry-picking mechanism allows users to select records from another MISP instance to be pulled to the local MISP instance. The MISP platform allows defining the distribution of the CTI records among organisations only, a community only, connected communities, and all sharing levels. MISP supports export of records and attributes in different formats (e.g., OpenIOC, CVS, STIX in XML, and JSON) to allow integration with other tools. Furthermore, it is possible to export signatures for IDS including Bro[12], Suricata[13], and Snort[14].

MISP is not only a software tool but also a series of data models created by the MISP community. MISP includes practical and straightforward information-sharing format expressed in JSON, which is the core format for the MISP platform itself. Moreover, the MISP format is described as an RFC draft. Its concept is based on objects, attributes and taxonomies. MISP attributes contain the pieces of data themselves. MISP attributes are of various categories and types, e.g. an attribute of type bank-account-nr belongs to financial fraud category. There are 14 categories and more than 150 default types. MISP objects allow building a collection of attributes. The objects are defined by a template that enumerates a set of attributes in the object. MISP also includes various existing taxonomies to classify events and attributes, such as CSIRTs/CERTs classifications, national classifications or threat model classification.

[12] https://zeek.org/
[13] https://suricata-ids.org/
[14] https://www.snort.org/

2.4 Actionable CTI

According to 1,200 IT and IT security practitioners surveyed in the United States and EMEA in 2017, the consumption and exchange of threat intelligence had increased significantly since 2015. However, most respondents were not satisfied with the exchange and use of threat intelligence. The inability to be actionable, timely and accurate was the most common complaint about threat intelligence [8].

Receiving and submitting information about vulnerabilities requires several processes before CTI can be called actionable. ENISA defines actionable CTI that fulfils five criteria: relevance, timeliness, accuracy, completeness, and ingestibility [9].

The research in [10] identified four stakeholders who work with CTI, namely, high-level executives, threat managers, threat analysts, and incident response teams. CTI data quality may differ by sharing stakeholder or source. The Quality may be evaluated by the correctness, relevance, timeliness, usefulness, and uniqueness [11]. Furthermore, a member of the CTI sharing community who has always shared useful and timely information may be labelled as a quality stakeholder [12].

Moreover, the threat environment changes quickly and thus, CTI must be acted upon quickly. The importance of sharing quickly can be observed when the value of CTI goes to zero in days or even hours [1]. As shown in previous research, 60% of malicious domains have a life span of one hour or less [13]. Timeliness does not only focus on age, but also on the frequency of updates to threat activities, changes, or evolution in capability or infrastructure [14].

From another point of view, organisations must prioritise the privacy of clients by sharing CTI only with trusted stakeholders or anonymise the content. Several matrices were developed to anonymise the content of shared information such as k-Anonymity [15], l-Diversity [16], t-Closeness [17], and others. Stakeholders are still reluctant to share information about breaches because of fear that it could damage their reputation, which is a valuable asset to protect [18]. Another aspect of anonymity is the encryption of CTI when shared between stakeholders. A Man-in-the-Middle attack could intercept the shared information. A protocol for encrypting CTI called PRACIS was presented in [19]. PRACIS enables privacy-preserving data forwarding and aggregation for semi-trusted message-oriented middleware. The work in [20] presented an architecture to compute privacy risk scores over CTI. The research discusses the privacy risks of extracting personal information from threat intelligence reports. Both presented works may be merged to enhance privacy in a CTI program.

2.5 Existing frameworks

2.5.1 NIST Cybersecurity Framework

National Institute of Standards and Technology (NIST) Cyber Security Framework is organised around 5 high-level domains:

- Identify
- Protect
- Detect
- Respond
- Recover

The NIST Cybersecurity Framework (see Figure 2.1) is a voluntary framework intended for critical infrastructure organisations to manage and mitigate cybersecurity risks based on standards, good practices and guidelines. It is a checklist that was built by considering numerous security professionals' experience and expertise.

The document is amended regularly (the last version is 1.1 from April 2018) in order to respond to non-critical infrastructure organisations. The framework helps individuals to take the appropriate decision and help with the communication about security measures, risks.

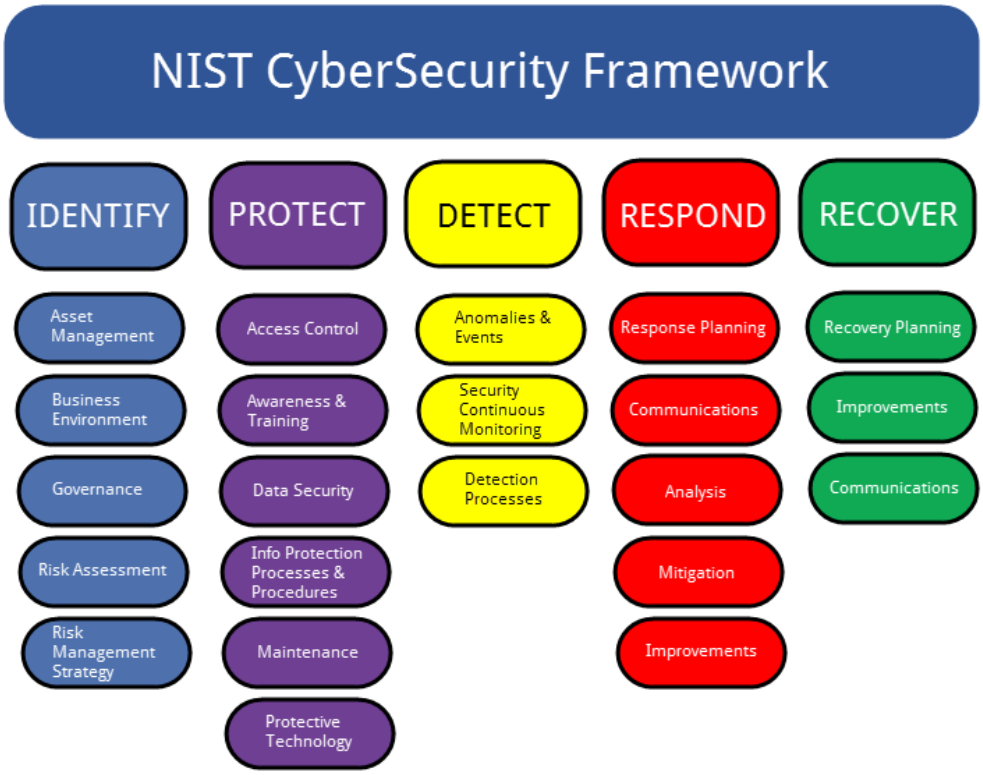

Figure 2.1: NIST Cybersecurity Framework

Identify. In order to manage cybersecurity risks associated with systems, assets, data and capabilities, it is essential to map the environment. In order to comply with this part of the framework there must be full visibility of the digital and physical assets, the interconnection between them as well as their roles and responsibilities. A proper understanding of the exposure to risks is crucial together with enforcement of the policies to manage those risks.

Protect. An organisation must develop and implement mechanisms to contain or at least limit the impact of a potential cybersecurity event. The organisation must enforce an access control mechanism to its digital and physical assets. The organisation must also provide awareness, education and training and put processes into place to secure data, maintain baselines of network configuration and operations to repair system components on time and deploy protective technology to ensure cyber resilience.

Detect. An organisation must implement measures to identify cybersecurity events swiftly. Monitoring solutions are required to detect anomalous activities and threats. The organisation must have complete visibility into its network to anticipate and prevent a cybersecurity incident. This visibility will allow having all the intelligence needed at hand to respond to the incident and make the hunting team very effective.

Respond. In the case of a cybersecurity incident, organisations must have the ability to contain it in order to minimise its impact. To comply, the organisation must implement a response plan, define an appropriate communication among the appropriate parties, collect and analyse all the information around and about the event. The response plan must plan activities to eradicate the incident and create a feedback loop in order to revise the future response strategies.

Recover. Organisations must implement effective activities to restore any capabilities or services that were impaired due to a cybersecurity event. The organisation must have a recovery plan in place in order to be able to coordinate restoration activities with external parties and incorporate lessons learned into their updated recovery strategy. Defining a prioritised list of action points which can be used to undertake recovery activity is critical for a timely recovery.

2.5.2 NATO AC/35-D/1017

The NATO security risk assessment guidelines are the guidance document for risk assessment and risk management for NATO systems. As such, it can serve as an example of risk management for military systems. However, it has to be noted that many NATO nations have published separate risk management guidelines and directives which can differ from the NATO guidance.

The document defines the role of risk assessment as a process of identifying security risks, determining their magnitude and identifying areas needing safeguards or countermeasures. It also provides extensive guidance on how risk assessment fits within the system development life cycle and how it needs to be executed. Note that the document does not prescribe a specific risk management method or tool (such as EBIOS), which means it is possible to apply an existing methodology and tool, as long as mapping can be provided from the external method to the guidance. Such a mapping is provided in the table below and can be used in interaction with the customers.

2.5.3 EBIOS

The EBIOS method was created in 1995 by the DCSSI (Direction Centrale de la Sécurité des Systèmes d'information) a government entity attached to the French Prime Minister. DCSSI is now called ANSSI. New versions were published successively in 2004 and 2010. The newest version includes compatibility towards international standards, in particular, ISO 31000 and ISO/IEC 27005.

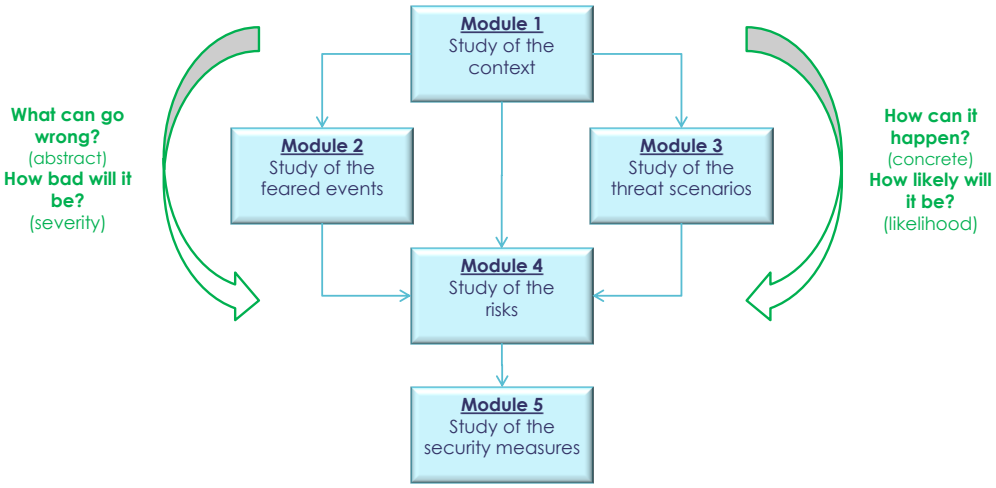

Figure 2.2: The EBIOS risk management rocess

The EBIOS method proposes five modules decomposed in activities, themselves decomposed in actions. The EBIOS activities and actions can easily be mapped to the risk management activities proposed in this guide.

Compared to ISO/IEC 27005, the EBIOS method mandates a clear cut between, on the one hand, primary assets and feared events, and on the other hand, supporting assets, and threat scenarios.

2.5.4 MITRE att&ck

Att&ck by the MITRE provides cartography of actions undertaken by the attacker once in the information system. MITRE ATT&CK is fully accessible and can be a starting point to elaborate a relevant threat intelligence approach. This framework can also be used to assess detection and defence mechanisms.

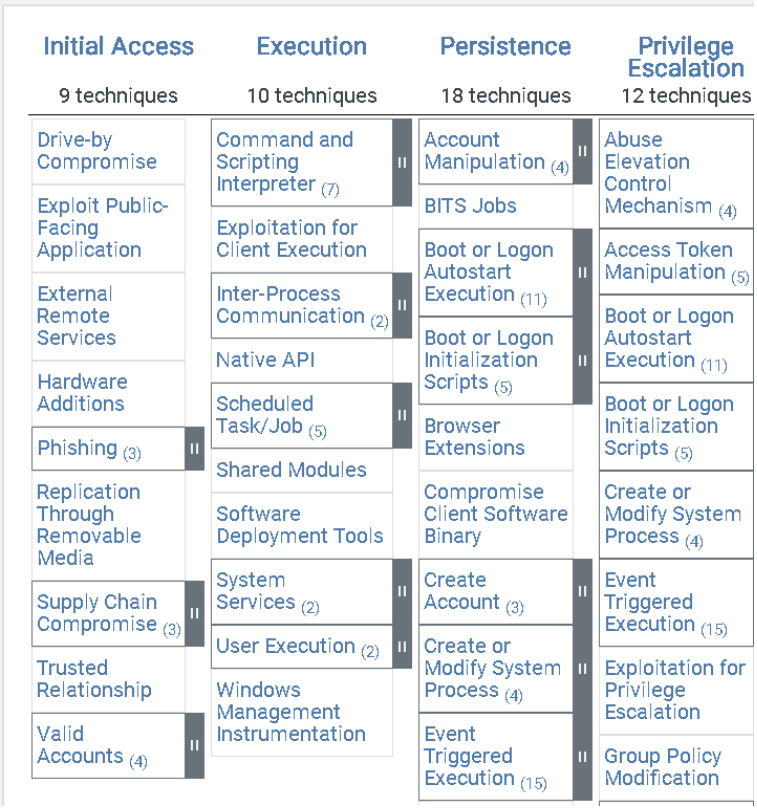

Figure 2.3: Part of MITRE Att&ck matrix

MITRE Att&ck is based on a matrix. The matrix provides an easy way to translate tactics and techniques of an attacker and extract the relevant intelligence in order to transform this raw data into actual actionable information.

Att&ck is very flexible; it can be used to

- visualise and assess available defence,
  - underlining attacks more suitable for the corresponding infrastructure,
- enhance the detection capabilities,
  - simply translating some of the rules of the framework into IDS/IPS rules,
- communicate,
  - the matrix is a user friendly and visual way of doing some cybersecurity education
- emulate an attacker behaviour,
  - creating multiple Red Team scenario.

2.5.5 ISO

ISO 22301. The international standard ISO 22301:2012 provides a best-practice framework for implementing an optimised BCMS (business continuity management system).

ISO/IEC 27001. ISO 27001 is the international standard that describes the requirements for an ISMS (information security management system). The standard framework is designed to help organisations manage their security practices in one place, consistently and cost-effectively.

ISO/IEC 27002. ISO 27002 is the companion standard for ISO 27001. Organisations cannot certify to ISO 27002, but the standard impacts ISO 27001 implementation by providing best practice guidance on applying the controls listed in Annex A of the standard.

ISO/IEC 27005. ISO/IEC 27005 provides guidelines for information security risk management. It supports the general concepts specified in ISO/IEC 27001 and is designed to assist the satisfactory implementation of information security based on a risk management approach.

Knowledge of the concepts, models, processes and terminologies described in ISO/IEC 27001 and ISO/IEC 27002 is important for a complete understanding of ISO/IEC 27005. The standard is applicable to all types of organisations (e.g. commercial enterprises, government agencies, non-profit organisations) which intend to manage risks that could compromise the organisation's information security.

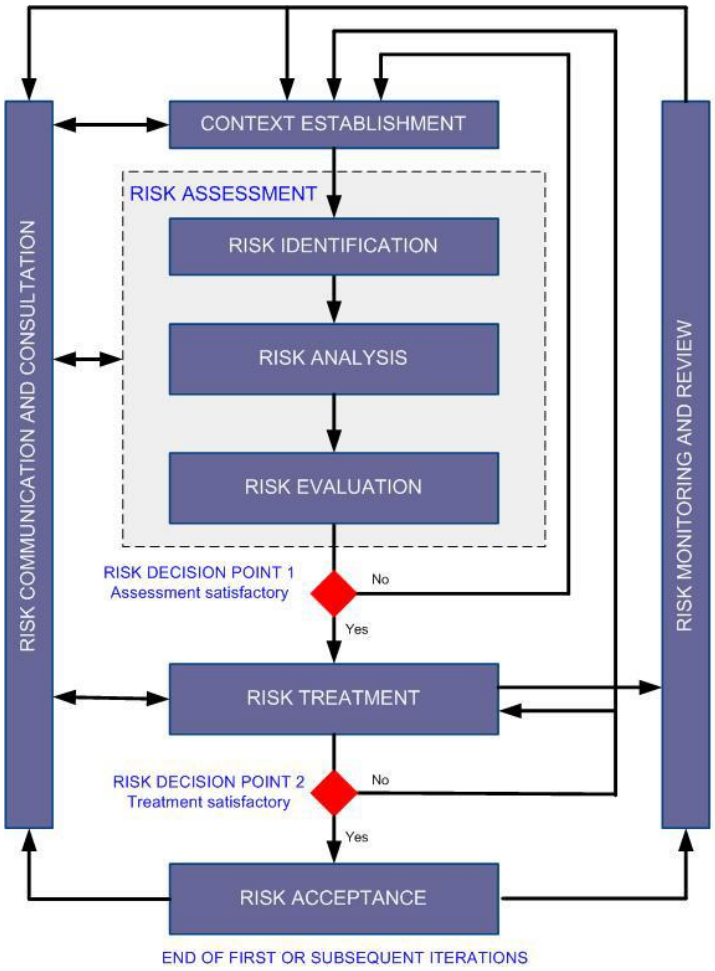

Figure 2.4: Risk assessment process according to ISO/IEC 27005

As shown above, the information security risk management process consists of context establishment (clause 7), risk assessment (clause 8), risk treatment (clause 9), risk acceptance (clause 10), risk communication and consultation (clause 11), and risk monitoring and review (clause 12).

ISO/IEC 27031. ISO 27031 provides a framework of methods and processes, improving an organisation's ICT readiness to ensure business continuity. Achieving compliance with ISO 27031 helps organisations understand the threats to ICT services, ensuring their safety in the event of an unplanned incident.

ISO/IEC 27032. ISO 27032 is the international standard offering guidance on cybersecurity management. It provides guidance on addressing a wide range of cybersecurity risks, including user endpoint security, network security, and critical infrastructure protection.

ISO/IEC 27701. ISO 27701 specifies the requirements for a PIMS (privacy information management system) based on the requirements of ISO 27001. It is extended by a set of privacy-specific requirements, control objectives and controls. Organisations that have implemented ISO 27001 will be able to use ISO 27701 to extend their security efforts to cover privacy management. This can help demonstrate compliance with data protection laws such as the CCPA and the EU GDPR.

2.5.6 ISA

One of the major activities of the Instrumentation, Systems, and Automation Society (ISA) is the development of standards for automation technologies. ISA's SP99 working group develops security standards for manufacturing and control systems, such as supervisory control and data acquisition (SCADA) systems and distributed control systems (DCS). An overview is available in the ISA Technical Report ANSI/ISA-TR99.00.01-2007, Security Technologies for Industrial Automation and Control Systems.

ISA's current cybersecurity standards are:

- ANSI/ISA-62443-1-1 (99.01.01)-2007 – Security for Industrial Automation and Control Systems Part 1: Terminology, Concepts, and Models
- ANSI/ISA-62443-2-1 (99.02.01)-2009 – Security for Industrial Automation and Control Systems: Establishing an Industrial Automation and Control Systems Security Program
- ANSI/ISA-TR62443-2-3-2015, Security for industrial automation and control systems – Part 2-3: Patch management in the IACS environment
- ANSI/ISA-62443-3-3 (99.03.03)-2013 – Security for Industrial Automation and Control Systems Part 3-3: System Security Requirements and Security levels

2.5.7 EPRI

The Electric Power Research Institute's Cyber Security Research Laboratory (CSRL) addresses the security issues of critical functions of electric utilities.

2.5.8 IEEE

The Institute of Electrical and Electronic Engineers publishes a number of standards on cybersecurity. Included are IEEE 1686-2013 – IEEE Standard for Intelligent Electronic Devices Cyber Security Capabilities; IEEE P1815 – Standard for Electric Power Systems Communications-Distributed Network Protocol (DNP3), sponsored by PE/PSCC – Power System Communications and Cybersecurity; and IEEE 1888.3-2013 – IEEE Standard for Ubiquitous Green Community Control Network: Security.

2.5.9 ITU

The ICT Security Standards Roadmap, published by the International Telecommunications Union (ITU), helps users to navigate the huge number of standards applicable to telecommunications, as well the organisations promulgating them.

Chapter 3 Use cases and requirements

The T-SHARK Programme is a practically oriented research and development package that revolves around the umbrella use case of election interference which is supported by individual subcases. The stage-gate process governs the development of subcases while the subcases are developed to provide practical requirements on the data model.

3.1 Methodology

This section aims to provide a short overview of the Stage Gates (SG) approach, used within T-SHARK programme (WP4). We describe the approach itself, applicability of SG and SG#1 event.

T4.1 organises the SG process within the scope of WP4. It was embedded in the SPARTA proposal. At the early stages of the project, it was planned that all the sub-cases could be reviewed in competitively at least twice. The actual structure and content of sub-cases proposed made some changes to the process. Sub-cases, developed within WP4, are very different and mainly focused on particular solutions. There is no possibility to compare them.

SG activities aim to govern Sub-cases for the better fit of T-SHARK concept of comprehensive cybersecurity and umbrella use case Elections interference. In addition, the SG process is structured in a way to facilitate the integration among Sub-cases and across other WPs.

Activities within T4.1 (Staging and pilot management) are focused around three main aspects:

- Governing of Sub-cases for a better fit to T-SHARK concept and umbrella use case (elections interference), aiming to provide sub-case owners with instruments enabling better fit and cross sub-case integration.
- Integrating developments outside SPARTA WP4 into SG process to enhance the concept of comprehensiveness, aiming to better fill-in the T-SHARK concept with more instruments ensuring the ability to demonstrate expected comprehensive threat management methodology.
- Organisation of SG events, aiming to provide an environment for Sub-case owners to present and AG members to review and be able to provide significant feedback.

The main focal point of activities is Stage Gates events. Aim of the SG events is to get the external point of view, questions and evaluation from Arbitrage Group (AG) members. It is not designed as a competition, as Sub-cases presented cannot be compared to each other, they have very different functionalities, targeted solution and other aspects.

3.2 Use cases

Umbrella use case: Election interference

T-SHARK Programme aims to develop and validate methodological, organisational and technological solutions extending cybersecurity towards the comprehensive organisation of security functions. The functions would focus more on threat prediction and full-spectrum cybersecurity awareness, providing high situational awareness, informing decision and policymakers on broad or long-term issues and providing a timely warning of threats. This focus requires to be scoped on the very complex, long term and beyond the technical interpretation of cybersecurity case. We selected Election interference umbrella case to be the main scenario for activities within the Programme. Elections interference, as an example of a strategic event, contains main features, allowing include, integrate, and validate the concept of comprehensive cybersecurity.

The umbrella case goes much further than the protection of the electronic environment as elections infrastructure itself. It includes physical and cognitive environments, which need to be connected by innovative methodological and technological solutions. Actors, having a significant role in the umbrella case, are many. Those include political parties, individual candidates, critical infrastructure, media, and public as the electorate.

Umbrella case covers the whole process of elections, starting with the announcement of candidates and ending with elections results. Thus, it can reflect long-term attacks and multi-stage exploits. Connecting cyber, info, kinetic events, understanding active actors behind those events allows building awareness, predictive abilities, and timely response for complex adversary actions. There is no single stand-alone system allowing the integration of all relevant methodologies and technological solutions. Umbrella case is used as a reference point, facilitating identification of relevant functionalities and value-added components in individual developments. Some developments in Sub-cases might be very technological, but by using umbrella case, they can be guided for improvement of relevant aspects, e.g. faster reaction, clustering of data, providing links with other events. An example of expected coverage of the umbrella case can be a recent event, related to COVID-19 and NATO. In this case, spoofed e-mails, dedicated accounts in social media, DDoS attacks, use of stolen identity and some other tools were used to destabilise decision making and create distrust.

Subcase 1: DDoS backscatter traffic detection

This subcase aims to propose an approach for DDoS backscatter traffic detection that can be utilised to improve the situational awareness for national CSIRT/CERTs and stratcom units, allowing them to correlate relevant information with other information sources and create a comprehensive view on large hybrid campaigns as early as possible. The current approach to detect backscatter traffic is to deploy honeypots. However, honeypots can observe only a limited range of the IP address space. In this subcase, we propose an approach that utilises the backbone network, hosting several /16 prefixes as an observation point for the backscatter traffic. In such case, it is not possible to use raw packet capture as a source of data, therefore flow collected from the edges of the backbone will be used to collect all backscatter traffic flowing through it. In such a setup, it is crucial to differentiate between backscatter, other malicious activities, misconfigurations and legitimate traffic. Machine Learning methods will be employed to train heuristics which will classify backscatter and non-backscatter traffic. The output of the ML classifier will be enriched with additional context data. The results should be shared within the cybersecurity community and should become inputs for comprehensive threat analysis.

Subcase 2: Detection of cyber and physical attacks on critical infrastructure across Europe

In this subcase, it is proposed an extension of mission-aware impact assessment models to incorporate information from multiple intrusion detection systems like Snort or Ossec, and other security and safety alarms such as firewalls, or physical sirens. Starting from the augmented VTAC, two additional dependency layers will be added: a physical layer to further include cyber-physical to the organisation layer that is the typical entry point for attacks occurring in SCADA networks. Physical layer will evaluate the status of physical components by calculating the impact of the organisation's physical assets. Moreover, since supervision of cybersecurity incidents and risks can no longer be done at the sole individual critical infrastructure level, a holistic dimension of the approach will be investigated. This will require the formal identification of interdependency links between critical infrastructure that can lead to the mechanism of propagation of an event within a critical infrastructure towards remote dependent critical infrastructures, as well as the method and associated algorithms for event propagation. The information thereby exchanged will feed a local prediction engine, to allow the analysis of the current and upcoming situation based on local data and data from the critical infrastructure ecosystem. AI technologies need to be completed with privacy-preserving tools such as Homomorphic Encryption (HE), and Private Aggregation of

Teacher Ensembles (PATE) approaches, in order to be able to exploit confidential data all along the lifecycle of AI methods, with a focus on the learning step.

Subcase 3: A specialized Virtual Control Room for protection of a critical infrastructure

This subcase aims to propose a Virtual Control Room that provides situational awareness for the cyber/ physical protection of critical infrastructures, providing operators with a virtual interface. The prototype will be demonstrated using assets that need to be protected. Such assets could be Leonardo's High-Performance Computer (HPC) and the Video Control System of Leonardo Chieti premises. The prototype will be demonstrated by considering as critical infrastructure the Leonardo's High-Performance Computer (HPC) and related assets, with connection to the video camera system present at Chieti premises. The core system is based on an OSINT platform that will try to identify vulnerabilities and malware related to the protected critical asset. The information will be captured in real-time and it will be available to the operator in the virtual environment for analysis. Furthermore, information on the safety status of protected assets and information from SCADA control systems will be shown in real-time, through virtual video walls. In particular, it will be possible to monitor all the alerts coming from the perimeter security systems (Firewalls, SIEM). A Decision Support function will be used to propose operators actions to manage Cyber-attacks/Physical intrusions. To summarise, the main capabilities of the Visual Control Room will cover the mapping and monitoring of the situations, comprehension of the situation, projection the effects that different actions may have and finally, support on the decisions and the actions that must be taken.

Subcase 4: Mapping of Future Events for Prediction of Cyber and Information Threats

This subcase elaborates on analysing possibilities, provided by the mapping of future events for prediction of cyber and information threats. It aims to develop a continually updated list of future events and their respective categories within the international, regional and domestic contexts of the EU members states together with augmented threat information (potential impact size, adversary's intention or similar). The database that will be produced may be used in areas such as strategic intelligence, predictive policing, and environmental protection. The subcase also aims to conduct an assessment of the tools which are used for the monitoring and mapping of such events. The results of the subcase analysis (IT system prototype, algorithms, descriptions/visualisation, aggregated/ structured database and methodology) could be later extended to other areas and integrated with other threat intelligence platforms. It will enable better threat understanding, from the current investigative-level definition, up to strategic considerations on current, future and down to real-time events handling and prevention.

Subcase 5: Modern Approach to Malware Analysis Automation

The overall objectives of the subcase is the development of automated tools support malware analysis on all stages, tracking the development of malware families and understanding the modus operandi of actors behind them. This will be achieved through the development of solutions to support malware analysts in assessing the type and functionality of the investigated samples. Specifically, the selected approach focuses on the detection of similarities between malware codes on various levels: entire unpacked (de-obfuscated) samples, functions and basic blocks. The subcase includes multiple methods of comparison: lexical analysis of decompiled code, comparison of normalized disassembly representation, API usage, control-flow graphs and more. Beyond facilitating analysis of individual samples, it enables an improved situational awareness and prediction capabilities through analysis of the overall development trends in the many malware families that are monitored by CERT.PL and other researchers. The prototype will be integrated with the online malware analysis and information sharing service created by CERT.PL – mwdb.cert.pl – which will make results of the analyses available to the research community. The second type of integration will focus on popular tools for reverse-engineering, like IDA Pro and Ghidra, to annotate individual functions with additional metadata that significantly speed up the process of manual analysis.

Subcase 6: Network flow-based threat intelligence method for a visual analytics system

This subcase aims to design a network flow-based threat intelligence method for multidimensional visual analytics system. Network intrusion detection is one of the main problems in ensuring the security of modern computer networks, Wireless Sensor Networks (WSN) and Internet-of-Things (IoT). In order to develop efficient network intrusion detection methods, realistic and up-to-date network flow datasets are required. Despite several recent efforts, there is still a lack of real-world network-based datasets which can capture modern network traffic cases and provide examples of many different types of network attacks and intrusions. By continually refining network flow-based methods, threat-related data can be identified, which can be used to anticipate technical threats and generate proactive solutions. The ability to process and analyse huge volumes of structured and unstructured data means that intelligent cybersecurity technical threat identification visual analytics systems can identify connections among data instances and technical trends that would be impossible for an expert to detect.

Subcase 7: Threats and attacks analysis

The present sub-case is centred in supporting LEAs in their actions to investigate and prosecute criminal organisations for launching cyber-attacks, focusing on the search of similar behaviours from the tactics and strategies used in the different attacks (represented by ATT&CK). The aim is to automatically (or pseudo-automatically) find relationships between threats and attacks in order to carry out the process of attributing malicious actions to an organised group (criminal, terrorist, state). This can be achieved through detection of common origins between threats and attacks. In the sub-case, the following challenges are addressed: a) The exhaustive characterisation of relevant variables and factors, both general and exclusive to each domain, that proves to maximise the collection and analysis of information in threat intelligence strategies for active defence, b) Research of habitual techniques for the construction of behaviour models of attacks that have been detected and characterised (attack model), c) Implementation of automatic learning techniques and algorithms, whether supervised or unsupervised, to group similar threats/attacks, which could be indicative of coming from the same source, and for the definition of commitment indicators that refer to the groups determined in the previous point.

Subcase 8: Anticipation of the cyber-physical attack on Transport CII

This subcase scenario will try to benefit from new security challenges that arise from the digitalisation era. In this context, one of them is the case of roads as a critical transport infrastructure. The participants will conduct three main activities, which will define the sub-case demonstration stages: First, digitalisation of the key cyber-physical asset of the covered CII infrastructure will be instantiated in a secure and isolated environment, where automatic or human-driven tests shall allow discovering specific Cyber Threat Intelligence like potential attack surfaces, cyber-to-physical propagations and evaluation of simulation-driven if-then scenarios. At the second stage, the gathered information will serve for guiding the custom hardening of the operational environment, definition of high-level safety/security policies and cataloguing potential courses of action. Finally, based on the outputs of the previous stages, capabilities for facilitating the acquisition of situational awareness (e.g. human-centric visualisations, human-in-the-loop simulations) and its projection at different time horizons will be conducted, which shall support reactive/proactive decision-making (e.g. anticipation of next stages of cyber kill chains).

3.3 Summary of requirements

This section aims to provide an overview of security requirements that must be satisfied by the common data model within the T-SHARK program. In order to represent specific data or relations in organisations' use cases, the data model must allow extensions to the standard specification. Furthermore, the extended data model must be backwards-compatible with the standard specification, thus allowing third parties to work with the data model including the extended part,

while these parties are not aware of the extension specifications. Moreover, to gather information from various sources that use different methods for collection, storing, and analysis of information, the common data model must be standardised, thus providing a special provision for the widest possible adoption.

Considering that sharing information between organisations can potentially lead to sensitive information disclosure, it is important to control that the common data model will not contain any information that can disclose an entity that suffers from the cyberattack. On the other hand, it is important to ensure that data is exchanged according to an expected schema that can be understood by other entities. Hence, the data model must be capable of being automatically checked against grammatical and construction rules and may be declared as valid or non-valid. This requirement influences the sharing capabilities of the common data model allowing entities to process data with their algorithms. Moreover, the data model also has to be easily understandable both by humans and machines to reduce the sharing and analysis processes.

Important requirements related to security purposes must be satisfied before, during, and after sharing information. Organisations may define rules and security policies for regulating access and usage of their data. Hence, the data model must allow the data producers to incorporate their security and privacy preferences directly in the data. Since each piece of data could have different security and privacy constraints, the data model must define proper fields and a proper format to host such information, because these constraints must be embedded directly in the data. Furthermore, to allow entities controlling the usage of their data that they are sharing, the common data model must include proper fields specifying security constraints including obligations. Hence, obligations will allow data producers to define necessary action to be enforced before, during, and after usage of the data. For instance, obligations may specify that whenever the data is accessed, the system must send the relevant message to the data producer or all entities that want to use data must accept a disclaimer.

A set of important requirements related to the analysis of threat information includes file properties, malware configurations, malware similarity, and malware clusters. Hence, the data model must include attributes for describing the results of static malware analysis. Furthermore, the data model must include proper fields to express static configuration extracted from malware samples and dynamic configuration obtained from command and control servers. Moreover, the data model must provide necessary fields to specify the numeric value for the similarity between malware samples as well as belongingness of a single malware to a cluster or clusters of malware.

The complete list of requirements, together with their description, is attached in Annex-A.

Chapter 4 Strategic analysis

The main purpose of our strategic analysis is to analyse the assumptions, limitations and challenges inherent in the SPARTA Cybersecurity Threat Intelligence Common Data Model as well as to make an informed decision what existing toolset to use to build the data model.

4.1 SWOT analysis

The conducted analysis embraced the SWOT Analysis methodology, which entailed a joint analytical action between SPARTA's partners, and defined a common discussion pool where the most outstanding key points of the proposal regarding the contributions and gaps of the state-of-the-art model were reviewed. The rest of this section presents the SWOTs analytical method and how it was instantiated in the scope of the SPARTA Cybersecurity Threat Intelligence Common Data Model, the adopted consultations and polling strategies, and the merged results. The Annex-B of this deliverable presents the separated studies internally conducted by each committed partner.

4.1.1 Background

The Strengths, Weaknesses, Opportunities and Threats analysis (SWOT) evaluates the internal strengths and weaknesses, and the external opportunities and threats in a competitive environment. The internal analysis is used to identify resources, capabilities, core competencies, and competitive advantages inherent to the methodology under review. The external analysis identifies market opportunities and threats by looking at competitors' resources, the industry environment, and the general environment. A SWOT analysis entails a widely applied and accepted methodology for strategic planning and management in organisations, which may highlight the main decision criteria, its factors and their relationships with the environment. Strengths and weaknesses are frequently internally-related, while opportunities and threats commonly focus on the external environment. The name is an acronym for the four parameters the technique examines:

- **Strengths**: characteristics of the object of study that gives it an advantage over others.
- **Weaknesses**: characteristics of the object of study that places it at a disadvantage relative to others.
- **Opportunities**: elements in the environment that the object of study could exploit to its advantage.
- **Threats**: elements in the environment that could cause trouble for the object of study.

The results are often graphically represented as a 2x2 matrix (see Figure 4.1), where the first column (Strengths, Opportunities) highlights the pros of the decision against the highlights in the second column (Weaknesses, Threats). On the other hand, the first row indicates the internal considerations (Strengths, Weaknesses) against the external, the latter being illustrated in the second row (Opportunities, Threats).

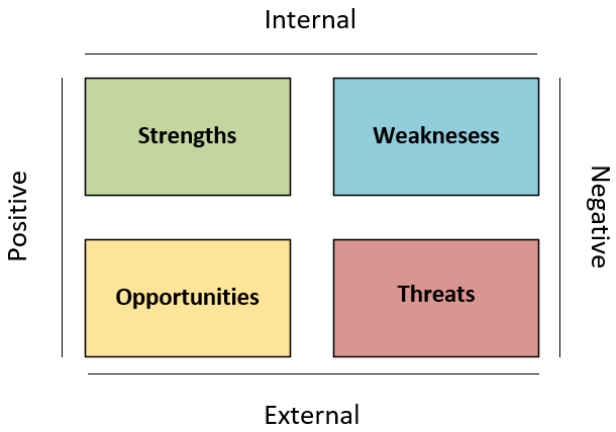

Figure 4.1: Graphical representation of the SWOT analysis results

The SWOT analysis typically entails the participation of 8-12 subject matter experts, being the heterogeneity of this advisory team critical for the proper interoperability of the results. The coordination of the exercise must preliminarily define a clear, one-sentence objective for the analysis; and all the experts must be aware of its context. In a brainstorming step, the experts will provide their vision concerning the SWOT properties; in a first round, the internal ones (Strength, Weaknesses) are typically covered and based on them, the external ones (Opportunities, Threats) are proposed. The results will be pulled, and the most significant and repeated indications will be prioritised. Their assessment and the context of the analysis will allow for guiding the next decision steps.

4.1.2 SWOT of CTI common data model

Within the T-SHARK Programme inside SPARTA framework the objective of the SWOT analysis is to produce a weighted merged SWOT matrix of possible candidates and components, for a new CTI Common Data Model. With this in mind, a clear objective has been presented to the participants:

> The Analytical study must identify the position of a hypothetical model built on the requirements adopted by the SPARTA CTI Common Data Model in terms of Strengths, Weaknesses, Opportunities and Threats, taking into account the existing features promoted by the related models in state of the art, their gaps and challenges.

The implementation methodology relied on the following considerations:

- The exercise must involve the wider and more heterogeneous group of experts, preliminarily polling the definition of SPARTA's beneficiary and associated partners involved (or not) in requirement analysis of the SPARTA CTI Common Data Model.
- The adopted method must be expandable and flexible enough for incorporating further feedback from other partners, stakeholders and cybersecurity practitioners. Therefore, the conclusions presented in the current version of this deliverable represent a snapshot of the study, which will evolve as new points of view are incorporated.
- The consultation method must be flexible enough to support remote contributions. Each participant has a different agenda (and even more so in times of a global pandemic), so it is assumed that it is not possible to gather a large number of partners during long time windows in which on-site SWOT analysis are usually carried out.
- The information that contextualised each analytical deduction regarding the state of the art (existing related data models) and the contributor is relevant. They must be present as metadata in order to facilitate further one-to-one comparisons.

Based on these grounds, we have adopted and enforced a distributed SWOT analysis development methodology. The analysis has been coordinated by IND which followed the following steps:

SPARTA Distributed SWOT analysis workflow:

1. The analysis coordinator (IND) notified the problem statement to the potentially committed partners, clearly indicating the purpose of the analysis, its modus operandi and responding suggestions/questions from the T-SHARK programme members.
2. An excel template was created and distributed to the participants. It served as a common register of the Strengths, Weaknesses, Opportunities, and Threats that each experts' team identified.
3. Each participant conducted an internal SWOT analysis, on which its cybersecurity and CTI experts: 1) highlighted the internal properties (Strength, Weaknesses) and based on them, 2) reviewed the external properties (Opportunities, Threats).
4. Each participant reported the internal analysis to the coordinator, which during a certain time period collected the different partner-side points of view.
5. The results were pulled in a unique SWOT analysis layout, indicating the most commented aspects, points of agreement and potential divergences.
6. From them, the coordinator raised overall conclusions, which are presented as the final results of the study.

Figure 4.2: Templates for distributed SWOT analysis

The template distributed to the participants was an excel file that requested the inputs illustrated in Figure 4.2. A table needed to be filled, which associated each proposed Strength, Weakness, Opportunity or Threat with a related data model (i.e., those being preliminarily described at the

above sections of this deliverable). Hence the reported observations were clustered by reference models, and their order of appearance.

Participants. The analysis gathered the perspective of seven T-SHARK programme partners: CESNET, EUT, IND, KEMEA, LEO, NASK and THALES; all of them are experts in the current Cyber Threat Intelligence landscape and its trends, and are perfectly aware of the requirements of the SPARTA CTI Common Data Model; further, some of them participated in the requirement analysis actions. The participation combined both industry (IND, LEO, THALES) and Research and Technological Development Organisations (CESNET, EUT, KEMEA, NASK), thus covering the full spectrum of business interests.

Their related background and suitability to participate as experts in the distributed SWOT analysis study are briefly justified in Table 4.1.

Table 4.1: Research and industrial entities expert in Cyber Threat Intelligence

Organisation	Description
CESNET	Cesnet z.s.p.o. (CESNET, Czech Republic). Association of legal entities, was held in 1996 by universities of the Czech Republic and the Czech Academy of Sciences. CESNET is the national e-infrastructure for science, research and education. It operates large backbone network infrastructures and provides services and connectivity to universities, campuses, research centres, schools, hospitals and selected government bodies. At the same time, CESNET is a research organisation with a long-term role as a technology innovator capable of transferring its research results into the commercial environment either by spin-offs or via its industry partners. CESNET research background ranges from network monitoring, traffic analysis, attack mitigation, incident handling, security event sharing to digital forensics. CESNET has a strong cooperation with Government CERT - GovCERT.CZ - operated by National Cyber and Information Security Agency of the Czech Republic (NUKIB) as well as with CSIRT.CZ the National CSIRT of the Czech Republic. From the international perspective, CESNET is a part of GEANT community as well as it participates in various European research projects.
EUT	Fundacio Eurecat (EUT, Spain). Eurecat is the leading Technology Centre in Catalonia, and the second-largest private research organisation in Southern Europe. The IT Security Unit includes ten highly skilled professionals: 3 PhDs in computer sciences, telecommunications, electronics, mathematics, security engineers experts in ethical hacking, working on the following research lines: 1) Cybercrime (detection and mitigation, identification of patterns and irregularities, the federation of information, deep web etc.; 2) Digital identity (privacy, anonymisation, authentication, etc.); 3) Distributed security (Internet of Things, Cloud Computing); and 4) Security in mobile platforms (Android, iOS).
IND	INDRA Sistemas SA (IND, Spain). Indra is one of the main global consulting and technology companies and the technology partner for core business operations of its client's businesses throughout the world. Through its Digital Labs division, Indra provides a response to the challenges of digital transformation. Digital Labs is comprised of the Indra experts in Big Data & Analytics, Cyber-security, Mobility, Smart Cities and IoT, amongst others. Its cyber defence and cybersecurity branches include among others: 1) PKI and Digital Identity, covering e-identity solutions for computer and mobile environments; 2) Biometrics, covering solutions that integrate biometric technologies for mobile, computer and kiosk (e.g., ABC) environments; 3) Cryptography and Blockchain, covering solutions and technologies with core cryptographic components and protocols (e.g., NetVote), as well as disruptive solutions around Blockchain technology; 4) Cyber

Organisation	Description
	Range, fully dedicated to cybersecurity training and experimentation solutions; and 5) Cyber Situational Awareness, focused on solutions and products for providing situational awareness in cyberspace.
KEMEA	Centre for Security Studies (KEMEA, Greece). KEMEA is a think tank on homeland security policies and an established research centre since 2005 (L. 3387/2005) within the Hellenic Ministry of Interior (former Ministry of Public Order and Citizen Protection), aiming to support security policy implementations in Greece, at a strategic level. The activity of KEMEA includes: a) the certification of practitioners in private security professions at the national level, b) research and development in the context of National and European projects in close cooperation with LEAs, working under the auspices of the Ministry of Interior and c) training of practitioners in new systems and technologies. The Centre also provides advisory and consulting services to the Ministry of Interior, as well as other Public and Private authorities, on safety and security issues.
LEO	Leonardo SPA (LEO, Italy). Leonardo is a global high-tech company and one of the key players in Aerospace, Defence and Security. The division which is carrying out Leonardo's activities in SPARTA is its Security & Information Systems Division. Through its “Cybersecurity and ICT Solutions” Line of business the Division delivers cybersecurity solutions, technologies and services that guarantee the security of data, networks and systems for critical infrastructures, government institutions, companies and individuals. In particular, Leonardo supports all activity phases: risk analysis, design and implementation of security architectures, training for prevention and management of incidents and disaster recovery. The Company is a reliable international partner for institutions such as the UK Ministry of Defence, the Italian Public Administration and NATO, thanks to its consolidated experience, as well as a team of specialists composed of analysts and defence experts. Leonardo is the only company outside the United States to develop and deliver a turn-key cybersecurity capability to NATO for the NCIRC FOC project that ensures the security of information and communications to around 70 NATO sites for a total of 70,000 users.
NASK	Naukowa I Akademicka siec Komputerowa Panstwowy instytut Badawczy (NASK, Poland). NASK is a research institute active in Poland. The specific groups that are taking part in the project are CERT (Computer Emergency Response Team) Polska and Network and Information Security Methods Team (NISM). CERT Polska was set up to handle Internet security incidents for the “.pl” constituency being operational since 1996. The research includes advanced data analysis methods, trust management, threat detection methods, virtualisation security, etc. NASK has contributed to many EU funded projects, under H2020, FP7, FP5 and the Safer Internet Action Plan, including coordination of the H2020 SISSDEN project. CERT Polska has significant practical experience in exchange and analysis of large amounts of security data. Additionally, CERT Polska brings experience with analysis and mitigation of botnet activity.
THALES	Thales Communications & Security SAS (TCS, France). Thales is one of the world-leading providers of mission-critical systems for security, defence and aerospace. TCS addresses every activity related to telecommunications: wireless communications, IP networks, satellite communication, network administration and security. TCS has a long experience in very large information systems and secure infrastructures for systems and networks, including the Internet and Intranets. TCS also develops a full range of telecommunication and cloud platforms and components, a range of high-performance security products and

Organisation	Description
	has a deep skill in secure telecommunications and information systems for public and governmental organisations, or emergency services. Advanced studies of TCS is made of a set of applied research laboratories working conjointly and involved in cutting-edge projects. ThereSIS laboratory representing advance studies in the SPARTA project covers four areas of expertise namely: AI & autonomy, data science, connectivity and cybersecurity.

4.1.3 SWOT conclusions and convergence

From the SWOT analysis and assessment of the contributions, “CCE” (which stands for Common Configuration Enumeration) has been mentioned several times. CCE is a methodology transferred to NIST management. Although widely used, CCE has yet to be integrated into a framework. Therefore, this might be a great opportunity.

On a note of interest, authoring a new member of the ISO27000 family has been suggested. However, it must be stressed that ISO27032 has been already authored, and maybe eventually our efforts could be in vain.

NIST Cyber Security Framework (from now on, ‘NIST’) is customizable to the nature of the organisation, therefore flexible and adaptive. NIST is very well structured, has straight guidelines and recommendations. However, there is a risk of bad integration, as NIST does not provide risk management guidance. Nevertheless, it is a very strong candidate.

NATO AC/35 is built mainly for military purposes; it is focused on risk management on military systems. However, several distinct risk management guidelines have been published by several NATO members. There is a “Risk of bad integration within an existing tool due to lack of existing mapping or difficulties into doing the mapping”. For all this, NATO AC/35 Strengths and Opportunities are overcome by Weaknesses and Threats, and it is most definitely discarded.

EBIOS is compliant with several international standards, being ISO standards among them. It is updated on a regular basis and has a strong community and institutional support. It has an exhaustive approach to risk analysis. Actually, its primary purpose is to evaluate Cyber Risks. This objective also originates one of its weaknesses: Cyber Risk Assessment is dependent on defining security requirements at the initial stages of a project and requires a significant level of expertise in security analysis.

MITRE Att&ck applies to all types of organisations, public or private, profit or non-profit, regardless of size or industry. It describes an attack from the attacker point of view, provides knowledge of the attacker and its profile. It includes controls and countermeasures for each of the tactics and strategies described. It provides a fully detailed approach based on a user-friendly matrix presentation, powerful and complete taxonomy on cyber threats and it is deep and rich in technical details. However, it is very complex, with a lot of information and a lot of attack patterns. Guidelines are too flexible and informal. The participation of security specialists is necessary for the definition of models of attack, tactics and strategies. Thus, MITRE Att&ck, while “only” describing attacks, is a very powerful companion in any CTI framework that talks about Advanced Persistent Threat (APT).

ISO27000 applies to all types of organisations, public or private, profit or non-profit, regardless of size or industry. It is a widely used Industry Standard. It allows security professional to check whether all information security gaps are covered in an Information Security Management System (ISMS). That is why any information security framework should be compliant with ISO2700 family. However, it is a high-level standard, enterprise-oriented and it addresses security from a Business point of view. Besides, applying ISO27000 standard and guidelines requires a specialist. In fact, there is a considerable risk of inadequately addressing information security while being compliant

with the standard, due to inadequate risk assessment or poor selection or implementation of security controls (as the organisation chooses the risk method and its security controls).

Mirkovic et Reiher has proposed a DDoS description methodology (taxonomy). It is highly specialized, covers all phases, including mitigation and recovery phase. It supports an interface to external sources.

STIX (STIX+TAXII+Cybox), has reached version 2. The goal of Cybox is to provide a common structure to represent cyber observables. It complements and links to different standards. The community widely recognises them as a complete and well-developed taxonomy for the representation of cyber incidents. The development follows a formalized and open process. STIX includes its own taxonomies, is flexible and has high granularity. It is a strong candidate and the most recognised standard. However, it suffers a usability threat: case defining on semantic level might be more resource hungry, as it is graph-based.

VERIS is a flexible data model for defining and exchanging incident information, developed by Verizon. Although the VERIS taxonomy can be used without the Verizon solution, it is often coupled with it in practical use (this could prove as a weakness). Although a private proposal, it has good recognition from the community.

OpenIOC is a flexible and extensible defined language for handling forensic information. It is useful for the representation of indicators of compromise, mostly used for detection in software. OpenIOC was not created for information exchange. Also, as a downside, OpenIOC is mainly used in MANDIANT's products.

AVOIDIT is a taxonomy that has not been updated since 2009, and has no strong community support.

CAPEC is another attack-focused data model. It is a well-documented methodology and there are several public data sharing databases. It can be integrated into STIX. As a downside, the participation of malware and attack analysis specialists is necessary to build the model of a given attack.

MISP, although originally designed for malware, is an open-source methodology. It is well structured, and can be used to share technical and non-technical information about malware samples, incidents, attacks and general intelligence. It possesses different levels and types to express threat, event or incident: Events, Objects, Object References, Tags, Sightings, MISP Galaxy. It is widely adopted and has several extensions, although these are perceived as too complex. MISP, as an EU based CTI, offers a unique opportunity to serve as a foundation for extension into a new CTI.

4.2 Expert workshop

4.2.1 Methodology

The goal is to achieve an expert decision on how to proceed when building a common data model for operational CTI. Throughout the preparation of the Workshop the core question that emerged and has still to be answered to was: shall something be created entirely from scratch, or shall an existing data model be re-used and possibly modified?

In order to make an informed choice, SMILE organised an expert workshop using SMILE's video-conference platform. The key experts involved in the discussion were Martin ZADNIK (CESNET), Evaldas BRUZE (L3CE), Sarunas GRIGALIUNAS (KTU), Jorge MAESTRE VIDAL (INDRA), Alexandre Dulaunoy (SMILE), Andras Iklody (SMILE) and Bertrand LATHOUD (SMILE). Bertrand Lathoud led the debate and ensured the agenda was followed.

Agenda

- Introduction round
- Question 1: Relevance of choice between the creation of a new data model and the modification of an existing one in the threat intelligence realm.
- Question 2: Problematic aspects of existing models such as STIX and analysis if they can become a “deal-breaker” or if they can be mitigated in an acceptably easy way.
- Question 3: If a new data model were to be created, do the merged requirements cover all needs? Would we be able to either add new ones, or simplify those already expressed as we are going to draft this new model?
- Question 4: Finalisation of the decision: what should be the practical decision process and who shall be involved.

4.2.2 Summary of the work session

The goal of the Expert Workshop was first to decide if the T-SHARK data model will be built on top of already existing models or if a new one will be created from scratch. The experience of MISP coverage of domains that are usually not directly dealt with by IT security specialists, such as Law Enforcement or Financial crime-related needs, showed that it is possible to have a fairly adaptable platform that is also widely shared and operates successfully. On the other hand, the lack of flexibility of the STIX process for adding new data objects illustrated the limitations of a standard that is too formally defined and cannot adapt fast enough to the ever-changing structure of the cyber threats realm. As a consequence, the experts came to the conclusion that the creation of a data model completely from scratch was unnecessary and not recommended. Different existing solutions have to be taken into account in order to create additional value to serve the T-SHARK programme values, which are primarily to go from reactive cybersecurity functions to pro-active cybersecurity functions. The idea is to perform comprehensive cybersecurity threat analysis, especially in predictive and analytic techniques. In order to handle properly complex threats, several domains beyond the cyber one, such as technical, social, informational and the physical domains have to be included in order to express relationships between various objects from these various domains. If it depends on an extremely slow and burdensome process for adding new objects describing emerging threats, it will not be of any help to operators in the field.

The experts agreed on the central requirement of multi-domain compatibility of the data model that will be created. This data model shall be flexible in the sense that it also can be used in other domains such as financial and legal, with a specific reference to the GDPR framework. It also has to be flexible enough to be later adapted for various domains, including heavily regulated ones, such as healthcare or military, for example. The idea is to start with a strong and extensible framework, which thereupon can be adapted in order to create additional value depending on how needs evolve. The creation of this layer will furthermore allow producing early results as it is directly related to sharing operational and technical information or data about the threats.

In the second phase, the experts analysed the gathered requirements for the data model and identified flexibility as one of the main requirements that have to be taken into account. Regarding the evaluation of the frequently used STIX model, some incompatibility issues were raised. The main disadvantage of the STIX model is, in particular, its lack of flexibility when dealing with data or information types not covered by the standard. STIX does not allow including anything that is not initially intended. This rigid focus would hamper the required multi-domain functionality, in particular for emerging threats, which are the target of the proactive analysis tools planned to be designed by the program. As the developers of MISP were involved in the creation of the STIX and TAXI model, they noted that there is no unique standard that could rule all use cases.

Further, the discussion revolved around the distinction between the creation of data taxonomy or ontology. MISP taxonomies allow for example to build specific data models. The MISP taxonomy and MISP galaxy are toolkits to create a corresponding model. The main critic point of the taxonomy is the lack of consistency in abstraction level, and the difficulty to perform transversal analysis of events expressed through different taxonomies. An ontology could in this way lead to added value in the sense that it would bring more formality and consistency, while staying generic

enough, which would lead to a better understanding by involved people that have no strong technical background, such as criminal analysts. Another issue that was raised is that taxonomies are very straight and that there would be less space for overlapping.

The main point is that one does not exclude the other. Thus, ontology can be built inside WP4 to describe use-cases in a formal way to allow people with different backgrounds to understand things in the same way. As a further step, the mapping, which would generate and describe overlapping, could be done using a platform such as MISP.

The primary outcome of the Workshop is that the decision part is not between data models but rather that these data models currently are often used by specific use-cases built by people working in different fields. It is very reactive and focused on solving very specific issues. It is not aimed at facilitating complex analysis but rather at sharing efficiently as much information as possible in order to get incidents under control quickly. Moreover, it embeds the sub-culture of the operators and investigators working in the field. These are elements that will not easily be perceived by the analyst working at a later stage on these data. Something more abstract, which would help to understand the information while not being acquainted with the sub-culture of the field would be beneficial for being able to perform proactive and complex analysis of the data related to several specific domains. The idea is to adopt a new standpoint without being constrained by the domain sub-culture of the existing data models. To summarise, the “generic data model” may actually be an ontology, which would be built on top of existing and successful models that were designed to solve specific classes of use cases.

4.2.3 Identified key requirements for the data model

Flexibility

Flexibility is an important characteristic, if not the most important one, as the goal of the data model is first to gather multi-domain information and intelligence. Second, a minimal system that can be extended on top is required for multi-domain functionality, such as sharing of cyber-physical information (i.e. influence or destabilisation campaigns). The model must be able to define incidents themselves, but as well to have the possibility to add information that is necessary for analytical processes to be applied later on.

Extensibility and adaptability

Another key requirement of the data model is its extensibility. The extension process of MISP is basically a construction mechanism that works with JSON objects and descriptions of the new objects. The only restriction is that these objects are a composite of different data objects like files, for example. A file is made up of certain types of hashes, filenames. Those individual data points that go into an object description, however, have to be in the default building block list of MISP. These types can easily be added in MISP and MISP has currently around 200 data types. The idea is here to use an existing model that is extensible, and to create additional value through the extension of this model. The evaluation and identification of the extensibility specifications of existing models will thus allow eliminating several models.

To illustrate the point, the way social media platforms, for example, use different APIs going from very basic to very complex ones holding several different attributes is particularly telling. These platforms use their own portals and defined native JSON formats, having attached raw incident information as JSON, including the original information sources. The unified data model requires a similar flexibility. Sources and data structures are changing in their architecture over time. Regarding social media intelligence, these changes are often recurring, and the T-Shark data model has to be able to be expanded and adapted over time in order to maintain its capability to analyse these feeds.

Flexibility, extensibility and adaptability are considered as must-have. This allows eliminating most of the present tools that have been assessed through the SWOT research led by INOV. They indeed represent very specific types of threat intelligence analysis capabilities, and were not

designed for a wide range and more importantly, a constantly evolving and broadening the spectrum of threats.

Formality

Regarding the ontology-based model, the idea is to take a domain, define attributes, identify what kind of attributes are covered, and then create a relationship with the data model. These relations could be used in order to build a data model that would be a transformation model between these domains. This transformation model would use XML or JSON and as a further step will be implemented through a MISP instance, to prove the model.

The main problem here is that by mapping, the data are translated into a common language first. The translation represents a risk of losing vital information for automated operations that can be performed if all the data from the different domains are expressed with one format only.

The ontology would create additional value in the sense that it would facilitate the construction of translators between taxonomies, and allow people with no strong technical background, to use the data model.

4.2.4 Next Steps

As mentioned, the data model has to be multi-domain proof; it's going to require several types of expertise and domains to be merged. The most important step towards a common data model for T-SHARK is to know how to build efficiently the list of the possible choices that fulfil the needs of the sub-cases, and to define these critical dimensions or criteria which are the most important for the specific requirements of a generic data model.

The working draft that shall lead to a decision will thus mainly revolve around two questions. The first question is to figure out which criteria and options will be selected and which will be eliminated. Based on the selected criteria, the options will be limited, and partners will have to describe additional factors that are needed to perform threat analysis in the field of their subcases. The second question that has to be resolved is to know what type of ontology will be built on top of the selected criteria and substructure.

Practically speaking, the sub-case owners will have to analyse the existing taxonomies through a platform such as MISP and identify whether the sub-cases can be expressed in those taxonomies. Subsequently, they find out which extensions are needed and finally, depending on the outcome of the ontology discussion, see how the taxonomies can be translated into one another or in a more generic one, a meta-taxonomy. This could result in several translations. A higher-level taxonomy, or meta-taxonomy, may be derived from the ontology research work. The meta-taxonomy would unify all the taxonomies into a widely accepted standard. This would enable all translations from specific into general taxonomies.

The sub-case owners could analyse the taxonomy trough MISP, identify those that are compatible with their cases, and then a composition approach could be applied. A composition would, in this case, be created from all the existing objects and attributes. This would lead to a new data model which would include all standards, frameworks and platforms already gathered through the research that has been carried out within the Sparta project. The final data model should be able to gather all kinds of attributes.

Chapter 5 Unification of data models

In this chapter, we propose a methodology to achieve unification of the data models in MISP and we demonstrate the methodology using election interference use case with selected subcases.

5.1 Methodology

The data model must evolve to reflect the needs of CTI sharing community which reacts to ever-evolving threats. Threat actors are coming up with new tactics, techniques, procedures and tools exploiting not only the technical domain but social, information and physical domains as well. At the same time, the cybersecurity community develops new concepts to improve the defender's capabilities, to render them broader and more effective.

The data model must be flexible enough to support these needs; at the same time, the data model should be consistent to support machine-readability. The process of enhancing the data model must be easy to perform as well as swift to enable the desired functionality as the need appears. On the other hand, if the process is too benevolent it may result in redundant and ambiguous representation which may be understood by a human analyst but renders the automated processing difficult and less reliable considering, especially, the cases when the machine should react adequately to the received information without human intervention.

Therefore we propose a methodology to support the way how new CTI models should be created in MISP to become commonly understandable. The methodology consists of several steps and follows the practical bottom-up approach outlined in this deliverable, i.e. it starts from the data output of each use cases, maps it to existing MISP models, identifies gaps and proposes extensions based on additional guidelines.

The first step of the methodology is to identify what data to share. This topic is elaborated in existing MISP document[15]. It is important to view the shared CTI from the perspective of the potential consumer (sharing partners), i.e. what data are relevant, how the received CTI can improve consumer's capabilities, what additional information the consumer needs to apply the received CTI. Additional supportive questions should relate to the stakeholders involved (e.g. threat actor, victim, sector, service), as well as dates and times, external publicly-available data that might be referenced and the description of the CTI itself which may become the source of categorisation of the given CTI.

The second step lists all the identified data points. For this deliverable, we utilise term data point to refer to an atomic piece of information, e.g. an IP address, a port number, a name. Each data point should be well described unless its meaning in the use case is not evident. While listing the data points, additional data points are usually identified which might be considered as an additional iteration through the first step. Subsequently, the data points are marked as either mandatory or optional. It is a good practise to consider multiple variants of available data and observe data points that always appear in all the variants and mark mandatory only a minimal subset of these points that provide such pieces of information without which it is not worth sharing the information at all. When a new variant of similar or additional data points appears, it is not necessary to infer a new model but to use the existing one. On the other hand, the fewer mandatory items, the more complex the processing of the model is.

The third step finds sets of data points that logically belong together and identify the necessary context. Typically, the data points in the set have implicit relationships between themselves and a common context. While these relationships are not explicitly expressed, they are inferred from the

[15] https://www.misp-project.org/best-practices-in-threat-intelligence.html

set by the CTI consumer and therefore, these relationships should be evident, and every consumer should infer the same relationships.

Any piece of contextual information is a significant contribution supporting the inference of the relationships or understanding of the data points. The contextual information relates either to the whole set or to individual data points. The contextual information typically captures characteristic, classification, categorisation or perspective. The contextual information may already be listed as a data point in which case it can be recognised as a qualitative measure rather than a quantitative measure. The universe of contextual information is rather limited and static in comparison to the quantitative data. The fourth step, therefore, consists of the identification of the contextual information from the existing data points and introducing additional contextual information.

The fifth step revolves around potential relationships that should be explicitly expressed in the data model. The explicitly expressed relationships are inevitable in cases when it is not straightforward to infer these relationships based on the context.

During the sixth step, the identified data points and groups, contextual information and relationships are mapped onto existing data concepts and models in MISP. While the previous steps are independent of the underlying technology and as such, they do not require knowledge of MISP, the sixth step requires an understanding of the MISP concepts and its available models. The MISP book[16] provides a user guide into MISP concepts such as events[17], attributes[18], objects[19], tags, taxonomies[20] and galaxies[21]. The MISP event represents a message that is shared by the platform. The MISP event contains attributes and objects which correspond to data points or logical group of data points respectively. The attributes, as well as the objects, can be tagged arbitrarily, but to keep the tag consistent there are various taxonomies which contain defined tags and their meaning. Galaxies represent more complex structured taxonomies. Tags, taxonomies and galaxies correspond to the contextual information. MISP also supports the expression of relationships by its dedicated relationship object.

We complement the mapping onto MISP with some best practise guidelines leading to a more consistent and thus more machine-readable CTI in MISP.

Guidelines

- Look for existing types and categories to represent a data point.

Typically, most of the data points in the list correspond to the actual data rather than the context and therefore will be represented by the MISP attributes. The MISP attribute must be of a certain type. MISP provides a large set of predefined types. The type belongs to one or more categories. The category expresses additional information about what the attribute refers to. E.g. type link belongs to category network activity and payload delivery. In the case of network activity, it means there was network activity/traffic towards this link observed while in the second case, it means that malware is delivered via this link.

- Use an object to express dedicated names of attributes.

An attribute in the MISP event has its type and category but does not have its dedicated name. In some cases the type might be sufficient to express the name (e.g. type ip-dst) while in other cases (e.g. type text) it is not clear what does the attribute represent. To this end, a comment might be assigned to an attribute. However, this hinders the machine readability. To assign names to the

[16] https://www.circl.lu/doc/misp/

[17] https://www.misp-standard.org/rfc/misp-standard-core.html#rfc.toc

[18] https://www.misp-standard.org/rfc/misp-standard-core.html#type

[19] https://www.misp-project.org/objects.pdf

[20] https://www.misp-project.org/taxonomies.html

[21] https://github.com/MISP/misp-galaxy

attributes an object needs to be created. Each attribute in the object has its dedicated name. The groups of data points identified in the third step are good candidates for the objects.

- Use an existing object, extend existing objects or create a new one

There are numerous already existing objects defined in MISP. For the sake of consistency, it is advisable to search through the existing MISP objects and to look for the most similar objects to the groups identified in the third step. The similarity captures the portion of the data points in the group corresponding to the attributes in the existing object.

In case there is a perfect fit or only limited fit, then it is straightforward, i.e. either to use the object as or to create a new dedicated object. When creating the new object it should be taken into account a broader scope of its utilisation by other similar use cases. Typically, the broader the scope, the fewer attributes will be marked as mandatory. See the second step and consider a reasonable trade-off.

In case there is a reasonable overlap with the existing object, i.e. the mandatory attributes can be filled and the object is only missing some additional optional attributes, then it is possible to extend this object. Moreover, it is possible to spread the new definition into other MISP communities by opening an issue or by pull request containing the updated object template into the MISP project[22] as MISP supports versioning of the objects. In case the intention is to keep the modification only local, i.e. for a given local community, then the modifications of the object might be more substantial.

- No nested objects

MISP (2.4) does not support nested objects (objects in objects). If such a feature is considered vital for the use case, a possible workaround is to utilise MISP relationship or to use a text attribute as the key reference to another MISP object id. The nesting of an object often captures an implicit relationship between the two objects. As a workaround it is possible to capture the relationship explicitly using the MISP relationship and keep the objects at the same level of the hierarchy, i.e. not nested. Please note that the MISP relationship is referred to as MISP relationship object which has its template (which might be extended) but it is not an explicit MISP object on its own.

- Use and extend taxonomies, galaxies

As identified by the fourth step, contextual information differs from the actual data. In MISP the contextual information should be expressed by tags, taxonomies and galaxies. While tags can carry any ad-hoc contextual information, they hinder the machine readability as they are not standardised. Therefore taxonomies and galaxies containing standardised tags should be prioritised. It is advisable to find corresponding taxonomies/galaxies and in case of missing tags to extend the existing ones or to create new ones. The new taxonomies, as well as galaxies, may be used in a local MISP instance only, but similarly to the objects, it is possible to publish them into other communities via the MISP project. Moreover, the upcoming version of MISP will allow editing galaxies on the fly which gives more flexibility what can be expressed.

- Link multiple low-level events

There may be multiple events published by different organisations. If there is a need to group these events, for example, to express that they are part of a complex campaign, then MISP extended event feature[23] could be used. At any time it is possible to create the high-level event describing the campaign from the high-level perspective (e.g. intentions, stakeholders) and then ask authors of the low-level events to extend the high-level event with the low-level events, i.e. to insert extended_uuid into low-level events which references the high-level event. See 5.2.

Template

[22] https://github.com/MISP/misp-objects#how-to-contribute-misp-objects
[23] https://www.misp-project.org/2018/04/19/Extended-Events-Feature.html

Collection
Use case
Describe the use case. Explain how sharing the output of the use case can contribute other partners in the community.
Data points
What data points should be shared? Which are optional? Which are contextual? Note: data points may be contextual and optional at the same time.
Sets
Do some data points logically belong together? If, so assign a name to the set and enumerate its data points.
Relationships
Capture the relationship between data points as well as groups. For example

Data point/set	*relationship*	*Data point/set*

Mapping
Types and categories
Assign types and categories to data points.
Objects
Are there existing objects to capture the proposed sets of data points? If yes, is it a perfect fit or shall there be an extension? If no, how the future object should look like?
Tags, taxonomies, galaxies
Are there relevant taxonomies or galaxies? If so explain their utilisation. If not, propose new one.
Relationships
Are all the relationships understandable implicitly or shall some of them be expressed explicitly.
Other
Other relevant features, such as references to other events. *Missing features in MISP itself.*

5.2 Data models

We apply the proposed methodology on several use cases related to election interference.

5.2.1 DDoS backscatter

Collection
Use case
The use case itself is described in Section 3.1.2. The output of the use case describes single vector volumetric DDoS attack targeting a single organisation. Sharing the knowledge about organisation under attack helps to understand the DDoS landscape, correlate the DDoS attacks with other attacks and follow the motivations and intentions of attackers.
Data points
• Start timestamp (mandatory) – observed start of the attack, • End timestamp – observed end of the attack,

- Victim IP address or prefix (mandatory) – the destination IP address or prefix that is under the flood of packets,
- Victim Port Number (optional) – in case the majority of DDoS traffic targets particular port number,
- Attacker IP addresses (optional) – source IP addresses if these are not spoofed
- Reflecting port number (optional) – in case of reflective attack, the port number of a service that is being misused for the reflection
- Reflecting IP addresses (optional) – in case of reflective attack, the IP addresses of reflectors
- Protocol (optional) – transport protocols (TCP, UDP, ICMP, …),
- Estimated number of DDoS packets (optional),
- Estimated number of DDoS flows (optional),
- Estimated number of DDoS bytes (optional),
- Estimated category of DDoS – classification of DDoS attack,
- Duration of service malfunction (optional) – duration for how long the service was not available for legitimate users or users experienced difficulties
- Sector (optional, context),
- DDoS type (optional, context)
- Organisation (optional),
- Address (optional),
- City (optional),
- Country (optional),
- Domain name of victim IP address (optional),
- Latitude (optional) – coordinate of victim,
- Longitude (optional) – coordinate of victim,
- Threat actor (optional) – who caused/hired the attack

Sets

Organisation = Sector, Name, Address, City, Country

DDoS = Start timestamp, End timestamp, Victim IP address or prefix, Victim port number, Attacker IP address, Reflecting port number, Reflecting IP addresses, Protocol, Estimated number of DDoS packets, Estimated number of DDoS flows, Estimated number of DDoS bytes, Estimated category of DDoS, Threat actor, Organisation, Latitude, Longitude

Relationships

Data point / Set	Relationship	Data point / Set
Victim IP address	belongs to	Organisation
Victim IP address	located at	Longitude
Victim IP address	located at	Latitude
Threat actor	Uses	DDoS
DDoS	targets	Organisation

Mapping

Types and categories

There already exists DDoS object in MISP, therefore, this step can be omitted as most of the

types and their categories is already defined within the DDoS object. The next step will elaborate the DDoS object.

Objects

Existing MISP DDoS object: https://www.misp-project.org/objects.html#_ddos

Some of the attributes of the proposed object already overlaps with the existing MISP DDoS object. The overlap of attributes, which are already included in existing MISP DDoS object, and proposed attributes is summarised in the following table:

Already defined attributes	Proposed attributes
domain-dst	victim_domain_name
dst-port	victim_port_number
first-seen	start_timestamp
ip-dst	victim_ip_address
ip-src	attacker_ip_address
last-seen	end_timestamp
Protocol	protocol

MISP DDoS object also defines attributes *total-bps* and *total-pps*, but these attributes carry units per second. Proposed attributes *number_of_ddos_packets* and *number_of_ddos_bytes* could be probably derived from these attributes and timestamp attributes *first-seen* and *last-seen*, but neither of these attributes is required, so derivation may not always be possible. For this reason, the attributes *number_of_ddos_packets* and *number_of_ddos_bytes* should be included in MISP DDoS object.

Timestamp attributes *first-seen* and *last-seen* are added to required attributes.

The next table summarises new attributes, which are missing right now in the current version of MISP DDoS object and are required for proper use of this object.

Custom name of an attribute	MISP attribute's category	MISP attribute's type	Attribute description	Correlation
reflecting-IP-address	Network activity	ip-src	In case of reflective attack, the IP addresses of reflectors	True
reflecting-port-number	Network activity	port	In case of reflective attack, the port number of a service that is being misused for the reflection	False
number_of_DDoS_packets	Other	counter	Estimated number of DDoS packets	False
number_of_DDoS_flows	Other	counter	Estimated	False

			number of DDoS flows	
number_of_DDoS_bytes	Other	counter	Estimated number of DDoS bytes	False
duration_of_service_malfunction	Other	counter	Duration (in seconds) for how long the service was not available for legitimate users or users experienced difficulties	False
latitude	Other	float	Latitude of victim IP address location	True
longitude	Other	float	Longitude of victim IP address location	True

Tags, taxonomies, galaxies

Existing MISP DDoS taxonomy: https://www.misp-project.org/taxonomies.html#_ddos_2

Existing MISP DDoS taxonomy offers predicate *type* for the description of DDoS attack. But these already defined types of attack can be extended by several useful values, namely *protocol-exploit-attack* and *malformed-packet-attack*.

DDoS attacks are also classified by other features, which should be added to existing taxonomy aswell. These features are:

Rate dynamics – dynamics of the attack

- Constant rate
- Fluctuating rate
- Increasing rate

Victim impact

- Disruptive – the goal of disruptive attacks is to deny the victim's service to its clients completely
  - It includes attacks, that denied service to more than 90% of users or more than 90% of an hour
- Degrading – the goal of degrading attacks is to consume some portion of a victim's resources, seriously degrading service to legitimate users
  - It includes attacks, that denied service to 60-90% of users or more than 60-90% of an hour
- serious - includes attacks, that denied service to 30-60% of users or more than 30-60% of an hour
- negligible - includes attacks, that denied service to 0-30% of users or more than 0-30% of an hour

- none – the attack had no impact on the victim

Victim type

- Application – targets a given application on the victim host
- Host – disables access to the victim machine
- Resource – attack targets critical resource in the victim network (e.g. DNS server)
- Network – attack consumes the incoming bandwidth of a victim network
- Infrastructure – attack targets some distributed service that is crucial for global Internet operation

For each of these features will be created new predicate with values specified in bullet points.

Relationships

All relationships are expressed well.

Other

Other relevant features, such as references to other events.
Missing features in MISP itself.

5.2.2 Twitter

Collection

Use case

This use case analysis twitter posts to derive their abusive content, misinformation or bias. Sharing this information helps to map activities of offenders as well as to correlate suspicious information contained in the tweets with other data. The shared data points describes tweet with multiple information about the tweet and also about user, who posted the tweet.

Data points

- *id* – unique identifier of the twitter post;
- *user-description* – description of the user;
- *user-location* – location of the user including country, region;
- *coordinates* – longitude, latitude;
- *truncated* – true/false
- *text* – twitter text;
- *hashtags* – hashtags in the twitter post;
- *urls* – links in the twitter post;
- *user-name* – defines the name of the user;
- *user-created* – date and time;
- *user-followers* – number of followers of the user;
- *id-str* – defines a specific identifier of the tweet;
- *created* – specifies when the post was created;
- *polarity* – specifies the negativity or positivity of the tweet on a -1 to 1 scale;
- *subjectivity* – defines how objective or subjective the tweet is, on a -1 to 1 scale;
- *quoted* – defines whether the twitter post is quoted;
- *quoted-text* – text of the quoted twitter post;
- *quoted-hashtags* – defines a list of quoted hashtags;
- *quoted-urls* – urls specified in the quoted tweet.

Sets

Twitter-account: *user-description, user-location, coordinates, user-name, user-followers,* and *user-created*.

Twitter-post: *text, hashtags, urls, id-str, created, polarity, subjectivity, text, truncated, quoted, quoted-text, quoted-hashtags*, and *quoted-urls*.

Relationships

Data point / Set	Relationship	Data point / Set
Twitter-post	belongs to	Twitter-account

Mapping

Types and categories

There already exist Twitter related objects in MISP, therefore this step can be omitted as most of the types and their categories are already defined within these objects. The next step will elaborate on these objects.

Objects

In MISP data model there are MISP objects to describe both Twitter-post and Twitter-account. However, the Twitter-post object does not have some attributes, so it will be extended with *polarity*, *subjectivity*, *quoted, quoted-text, quoted-hashtags*, and *quoted_urls* attributes. Hence, the existing MISP Twitter-post object will include these attributes. Furthermore, it is possible to extend the object with the *harassment-text* and *is-harassment* attributes, additionally to the existing *possibly-sensitive* attribute. The *harassment-text* attribute can specify whether the posted tweet has unacceptable content, while the is-harassment attribute can specify the numeric value using.

The MISP object **Twitter-post** will be extended with the following attributes:

Custom name of an attribute	MISP attribute's category	MISP attribute's type	Attribute description	Correlation
polarity	Other	text	Specifies the negativity or positivity of the tweet	False
subjectivity	Other	text	Defines how objective or subjective the tweet is	False
quoted	Social network	Boolean	Defines whether the twitter post is quoted	False
quoted-text	Social network	text	Text of the quoted twitter post	False
quoted-hashtags	Social network	text	Defines a list of quoted hashtags	False
quoted-urls	Social network	text	URLs specified in the quoted tweet	True
harassment-text	Other	text	Specifies the	False

			harassment part of the tweet	
is-harassment	Other	float	Defines whether the content of the tweet is acceptable on -1 to 1 scale	False
created	Other	datetime	Datetime of creation of the post (when it was posted)	True

Also MISP object Twitter-user will be extended with the attribute *created*, which describes datetime of user creation.

The MISP object **Twitter-user** will be extended with the following attribute:

Custom name of an attribute	**MISP attribute's category**	**MISP attribute's type**	**Attribute description**	**Correlation**
created	Other	datetime	Datetime of user creation	True

In fact, the *harassment_text* and *is_harassment* are similar to the polarity of the tweet. However, it can be used to specify harassment text and give the numeric value depending on the content. The *polarity* defines how positive or negative the context of the tweet. Values of polarity and subjectivity are calculated using the textblob NLP library. In fact, the polarity and subjectivity can be used together with other NLP functionalities to determine whether the tweet post has content for non-profit positive and negative advertisement of political party, politician. For example, the tweeter post might advertise voting for a specific political part or contain discredit information as a part of a campaign against the party or specific candidate.

Existing MISP Twitter-account object can be extended with an attribute that specifies a list of unique identifiers of MISP twitter-post objects instead of the attribute that specifies the total number of tweets posted by the user.

Tags, taxonomies, galaxies

Most of these Twitter posts will be an attempt at disinformation spreading or some kind of harassment. For these kinds of events there is taxonomy called *DRFLab-dichotomies-of-disinformation*. Mainly predicates *target-category* (specifies who the target of the disinformation is), *platform-social-media* (will be followed by value *twitter*) and *content-topic* (specifies the subject of the disinformation).

DRFLab-dichotomies-of-disinformation taxonomy: https://github.com/MISP/misp-taxonomies/blob/master/DFRLab-dichotomies-of-disinformation/machinetag.json

Relationships

All relationships are expressed well.

Other

Twitter is social media platform globally used mainly by political parties, influencers. Many big news companies take information from this platform and its posts. It can be then easily exploited for disinformation and harm some political party before the election. Therefore such twitter posts should be correlated with other events or linked to the more complex campaign.

5.2.3 Election interference

Collection
Use case
The purpose of this use case is to allow MISP users from different CTI organisation to share intel on Election related events, in a structured way.
Data points
• Date when the election interference occurred (mandatory) • Date of the targeted election (mandatory) • Name of the targeted election (mandatory) • Threat actor who is responsible for the interference (optional) • Threat actor motivation (optional) • Modus operandi (optional)
Sets
Election interference campaign = includes all of the above attributes
Relationships

Data point / Set	Relationship	Data point / Set
Threat actor	targets	Election events

Mapping
Types and categories

Custom name of an attribute	MISP attribute's category	MISP attribute's type	Attribute description	Correlation
Date	Other	datetime	Datetime when the interference campaign occured	True
election-event-date	Other	datetime	Datetime of the targeted election	True
election-event	Other	text	Name of the targeted election	True

threat-actor	Attribution	threat-actor	Name of the threat-actor who was involved in the interference	True
threat-actor-motivation	Other	text	Attacker's motivation, political views or ideology	True
modus-operandi	Other	text	The threat actor technics	True

Objects

In MISP data model there is no similar object as Election interference, so completely new MISP object will be created.

In this new object, threat-actor-motivation attribute will be compliant with MISP/vocabularies/threat-actor/cert-eu-motive while threat-actor-motivation will point out specific MISP/ttp-category.

Tags, taxonomies, galaxies

In MISP data model there is already defined Election guidelines galaxy. But this galaxy should not be used directly in this election interference event to prevent duplication of information, because individual subcases will use this galaxy. These used galaxies can be then easily displayed in this high level event.

Election guidelines galaxy: https://www.misp-project.org/galaxy.html#_election_guidelines

But the election interference event should use at least some tagging information, which will convey, that the event carries information about elections. From this perspective will be enough to use current-event:election tag, which will sufficiently classify the event and more detailed classification will be already included in concrete election interference subcases via Election guidelines galaxy.

Current event taxonomy: https://www.misp-project.org/taxonomies.html#_current_event
Other MISP galaxy vocabularies fit for description of:

- Threat actor
  misp-galaxy/vocabularies/common/threat-actor-type.json
- Threat actor motivation
  misp-galaxy/vocabularies/threat-actor/cert-eu-motive.json
  misp-galaxy/vocabularies/threat-actor/intended-effect.json
  misp-galaxy/vocabularies/threat-actor/motivation.json
- Modus operandi
  misp-galaxy/vocabularies/common/ttp-category.json
  misp-galaxy/vocabularies/common/ttp-type.json

Relationships

Relationship	Description	Format
targets	Governmental	MISP/Organisation

	institution	*Role* : Victim *type-of-organization* : *misp-galaxy:sector="Government, Administration"*
targets	Political Party	MISP/Organisation *Role* : Victim *type-of-organization* : *misp-galaxy:sector="Government, Administration"*
targets	Candidate or alternate	MISP/Person

Other

This event will carry important data about election interference itself and will be used as the main high-level event of whole election interference campaign and will reference to concrete interference subcases.

5.2.4 BP-IDS

Collection

Use case

Business Process-based Intrusion Detection System (BP-IDS) is a process monitoring solution that aims at the detection of incidents on technology-enabled infrastructures. It matches in real-time the activities detected in the executed business process with the specified business process and specified business rules. Whenever that executed process deviates from the specification, the activity is marked as a possible incident and the infrastructure administrator is notified in real-time by BP-IDS with the causes of that anomaly (traces, affected processes, etc.).

Example of the incident could be the service of supplying water to a city. A failure to measure the quality of the water-based on sensor readings would raise a FailedActivity. A failure open the valve on water delivery pipeline, would first raise a FailedActivity, and then a FailedProcess describing failure on the water delivery service.

Data points

- Engine specific information (mandatory)
- name of the core component
- IP address with port of the core component
- Detection sensor specific information (optional)
- name of the sensor
- IP address of the sensor
- sensor type – e.g. Network sensor
- alert type classification – can be either “Failed Activity” or “Failed Process” (mandatory)
- alert description - text created by BP-IDS with alert description (mandatory)
- additional data resource – optional link to more information about the alert (optional)
- occurrence time – time when the alert occurred (mandatory)
- detection time – time when the alert was detected (optional)
- source or destination host information (optional)
- ip of the host
- port number
- application – name of the host application
- role – role in the communication (source/destination)

Sets

BP-IDS Alert = engine specific information, detection sensor specific information, alert type classification, alert description, additional data resource, occurrence time, detection time
Host = source or destination host information (ip, port, application, role)

Relationships

Data point/Set	Relationship	Data point/Set
Source Host	Communicates with	Destination Host

Mapping

Types and categories

See the next step.

Objects

In the MISP data model there is no similar object as BP-IDS alert, but for Host object can be used *ip-port* MISP object. So one completely new MISP objects will be created and *ip-port* has to be updated just with *application* attribute. If the host's role is destination or source should be clear from attributes used (ip-src or ip-dst).

The BP-IDS-alert object will consist of the following attributes:

Custom name of an attribute	MISP attribute's category	MISP attribute's type	Attribute description	Correlation
engine-name	Internal reference	text	Name of the core component (engine)	False
engine-ip-port	Network activity	ip-src|port	IP address and port of the core component	True
detection-sensor-name	Internal reference	text	Name of the sensor component	False
detection-sensor-ip	Network-activity	ip-src	IP address of the detection	True

			sensor	
detection-sensor-type	Internal reference	text	Type of the sensor	False
alert-type	Network activity	text	Type of the alert	False
alert-description	Internal reference	text	Text created by BP-IDS with alert description	False
additional-data-resource	External analysis	link	Hyperlink for BP-IDS Monitor Web User Interface	True
occurrence-time	Other	datetime	Timestamp of alert creation	True
detection-time	Other	datetime	Timestamp of BP-IDS Sensor data collection	False

MISP *ip-port* object: https://www.misp-project.org/objects.html#_ip_port

Existing *ip-port* MISP object will be updated with following attribute:

Custom name of an attribute	**MISP attribute's category**	**MISP attribute's type**	**Attribute description**	**Correlation**
application	Internal reference	text	Name of the host application	False

				used	

Tags, taxonomies, galaxies
In MISP taxonomies there are two existing taxonomies to express IDS alert detection. One is ecsirt:intrustion-attempts="ids-alert", second one is rsit:intrusion-attempts="ids-alert". In this type of event there should be no other needed classification; at least it cannot be derived from information the alert carries. ECSIRT taxonomy: https://www.misp-project.org/taxonomies.html#_ecsirt RSIT taxonomy: https://www.misp-project.org/taxonomies.html#_rsit
Relationships
Other
With respect to election interference the BP-IDS events will be reporting incidents happening on critical infrastructures such as: power plant shutdown and water delivery problems such as, polluting water to compromise the safety of European citizens and shutdown distribution of water.

5.2.5 Malware

Collection
Use case
The subcase is described in the section 3.2 (Subcase 5: Modern Approach to Malware Analysis Automation). Only a subset of the data used within the subcase will be exported to MISP. Specifically, export to MISP will include elements that are relevant for providing context to existing events and enable correlation. Full malware datasets will not be replicated into MISP as purpose-built systems (for example MWDB) are better suited for storing a large amount of the subcase data.
Data points
• Labels / name of malware family of a given sample. • Cluster names that a given sample belongs to. • Other known samples that share similarity to the given sample. • Details on the type and degree of similarity (e.g. code reuse in a number of functions).
Sets
Data in the MWDB platform, results from Drakvuf-sandbox and results of similarity analysis.
Relationships
Relationship to other samples (see "Data points" above).

Mapping
Types and categories
n/a (see "Objects")
Objects
New object template: malware-similarity

Meta-category: file

Object attribute	MISP attribute's category	MISP attribute's type	Attribute description	Correlation	Multiple
original-sha256	External analysis	sha256	The SHA256 of the sample that is being analysed.	True	False
original-md5	External analysis	md5	The MD5 of the sample that is being analysed. (allows correlation of samples that do not have SHA256 available in MISP)	True	False
similar-sha256	External analysis	sha256	SHA256 hashes of sample that are similar to the one being analysed.	True	True
similar-md5	External analysis	md5	MD5 hashes of sample that are similar to the one being analysed. (allows correlation of samples that do not have SHA256 available in MISP)	True	True

cluster-label	External analysis	text	Names of clusters that the analysed sample belongs to. The names can correspond to malware family name or just provide an opaque unique identifier for correlation.	True	True
analysis	External analysis	comment	Additional information on the analysis for diagnostic purposes.	False	False

Tags, taxonomies, galaxies

Distribution level of individual attributes will be set to “inherit” (default) except for relationship with samples with limited visibility in

Additional details of the similarity between samples will be saved in comments (metadata in free-text) in the “similar-sha256” and “similar-md5” attributes.

mwdb taxonomy (Malware Database (mwdb) Taxonomy - Tags used across the platform); these tags will be automatically added to attributes referencing samples.

The following taxonomies are suggested for analysts contributing events representing results of investigations involving malware analysis, but their use is not strictly required by the subcase itself:

- TLP taxonomy
- Malware galaxy (Name of ATT&CK software, UUID: d752161c-78f6-11e7-a0ea-bfa79b407ce4)
- Malpedia galaxy (Malware galaxy based on Malpedia archive, UUID: 1d1c9af9-37fa-4deb-a928-f9b0abc7354a)

Relationships

Related files are linked automatically in MISP through the built-in correlation functionality that links attributes, and consequently events, that have the same type and value. “similar-sha256” and “similiar-md5” attributes will link to other samples in the system, even if they are not part of the “malware-similarity” objects or directly related to the subcase.

In the future, an advanced correlation feature in MISP can be extended to automatically correlate samples by looking up their similarity in the external tools provided by the subcase.

Other

The malware-similarity objects will be added to events when an analyst using MISP uses the

enrichment feature to get more information on malware hashes that are relevant to the investigation. Therefore the implementation of integration with MISP will focus on development of an enrichment module, that will be triggered manually or via the MISP API for attributes of interest.

Chapter 6 Demonstration

This section demonstrates the applicability of selected extensions of the MISP model in a practical deployment; in particular, we demonstrate the SPARTA T-SHARK programme umbrella use case of election interference.

6.1 Deployment

First, we deployed a MISP instance dedicated for SPARTA. To this end, we created a virtual server running Debian version 9 with a public IP address so that SPARTA partners can access it. The server has two vCPUs, 512 MB of memory and 20GB of disc capacity, which is sufficient for initial setup and we will add resources on-demand on the fly due to virtual environment.

The MISP instance was created using the official installation script[24] available from the MISP repository ("MISP core" and "MISP-modules" components were installed). After the script installed all dependencies and the MISP itself, we set up the admin account, generated a new PGP key-pair and imported it to MISP's keychain (it will be used to sign emails sent by the MISP instance), and we tested that emails sent from the server are successfully delivered.

Then we configured the MISP instance – the following parameters were changed from defaults:

- MISP.external_baseurl = https://misp-sparta.liberouter.org
- MISP.email = no-reply@misp-sparta.liberouter.org
- MISP.org = SPARTA
- MISP.contact = (admin's email address)
- MISP.footermidleft = This is an experimental instance of MISP for the SPARTA project.
- MISP.footermidright = Operated by CESNET.
- MISP.welcome_text_top = (empty)
- MISP.welcome_text_bottom = (empty)
- GnuPG.email = misp@misp-sparta.liberouter.org
- GnuPG password = (password of the private GnuPG key)
- SMIME.enabled = true

The SPARTA MISP instance is running and its web interface is available at:

https://misp-sparta.liberouter.org

The partners joined the SPARTA MISP instance using a standard procedure. An "organisation" was created in MISP for each partner by the SPARTA MISP administrator. Each partner selected one person to be the administrator of its organisation. This person sent an email with their SMIME certificate or PGP public key to the SPARTA MISP administrator, who then created the corresponding user account in MISP and set up its permissions. During the account creation, the MISP instance automatically generated a random password and sent it to the user via an encrypted email. These users then added additional users per their respective organisation as they saw fit.

24 https://misp.github.io/MISP/INSTALL.ubuntu1804/

6.2 Demonstration of election interference

The demonstration is based on a fictional scenario of election interference campaign. First, we describe the scenario of election interference and then we show how its certain parts reflect in the cyber threat intelligence using the proposed extended data models of MISP.

6.2.1 The fictional story of election interference

The Republic of Peripheral Europe is a small (still) democratic country with over 10,000,000 inhabitants with external EU borders. Thirty years ago, it emerged from the former socialist block of countries and joined the EU 15 years ago. The country has a vibrant political scene and holds elections every four years. Moreover, it is one of the most internet-connected countries in the world, where the population receives most of its information from internet newspapers and social networks.

Unfortunately, the political situation in the country escalates due to some recent social and economic events, and as it is usual in such times, extremist and radical parties can gain popularity. Elections are a very sensitive democratic process that can be an easy target for such groups.

This story presents a model scenario of a political activist group aka Radical Pirates that tries hard to interfere with the national election. This election is essential since the popularity of various radical extremist groups that proclaim autocratic regimes are increasing and it represents a severe threat to democracy. Radical Pirates group represents a majority of the extremist groups and therefore they are the most active to reverse the results in their favour at the expense of traditional democratic parties with the currently leading FreeWill party.

Radical Pirates want to draw the attention of the population and damage the reputation of FreeWill members at any cost. They know that many citizens regularly check the content of the FreeWill party website to read the latest news, updates of the legislative process regarding prepared laws, and long term strategy. Especially in the time of the upcoming election, it is essential for every political party to update websites with relevant information.

Since Radical Pirates want to discredit their main opponent, i.e. the FreeWill party, they plan to create a modified copy of the FreeWill web pages that can be hosted by anonymous webhosting service. Naturally, this web page with fake content must look legitimate and should be visited by as many voters as possible, so Radical Pirates buy several domain names that are remarkably similar to the original freewillparty.rpe domain name. They rely on the low attention of users that will not recognise easily that they access fake URL addresses. Additionally, dissemination of the fake pages is also supported by massive spam e-mails and social media posts with the hyperlinks pointing to the fake websites.

One of the polarising political topics of the last few years in RPE is ecology and energetics. FreeWill pushes for the quick end of coal power plants in favour of renewable sources (solar and wind) that are being built massively throughout the country thanks to generous state support. Their opponents criticise the plan. Among others, one of their arguments is the outdated power grid of the country that is (reportedly) not ready for several decentralised sources with fluctuating power output.

Radical Pirates cleverly misuses this against FreeWill. With the help of HackAllWorlds, they find a vulnerability in a control system used at several electrical substations. They manage to exploit it to disrupt power distribution in such a way that it causes a blackout affecting the whole region of the country for several hours. Moreover, they deliberately timed the attack to happen during a celebratory opening of the largest solar park in the country. Radical Pirates take care of spreading a word via trolls and fake accounts on social media connecting the blackout with the growth of renewable sources, which were claimed as highly unreliable.

FreeWill is very active on social media. In fact, it is the most common instant way to communicate with their supporters. FreeWill has one shared account for all popular services (such as Twitter, Facebook, Instagram), and a few selected trusted party members know the credentials to post relevant messages.

In this scope, Radical Pirates created another plan of how to attack FreeWill and alter public opinion by misusing the existing account of FreeWill.

Radical Pirates have good experience with spear phishing. They gathered the contact list of FreeWill members. They plan to infect devices of the FreeWill members to steal access credentials to social media accounts with specially crafted malware. HackAllWorlds group provides their malware code that scans victims' devices and searches for stored credentials, so the plan is to attach an infected file (with a fake marketing strategy) to the emails sent to FreeWill members. By the way, the email message urged everybody in FreeWill to check their password in the information system because of increased hacking activity of opponents and to study the attached marketing strategy. This warning and marketing strategy deceives most of the FreeWill members before they are warned their devices become infected and multiple access credentials to various accounts are disclosed to the Radical Pirates group.

Radical Pirates got another special weapon from HackAllWorlds in the form of the additional feature in the second malware sample. This malware was prepared to target potential voters of the democratic parties or still undecided citizens-voters. The behaviour of the malware was designed to wait on the background in the victim’s operating system and act whenever a user connects to any advertisement server. The aim is to affect the commercials and banners that are shown to a user. Their content blames democratical parties as a source of issues in the current society and proposes a different way that is represented by Radical Pirates.

Besides advertisement, the malware process is also able to modify the content of social media web pages - hide selected posts, users or pages, or inject fake posts, as instructed by the command&control server. The aim is to decrease the visibility of potentially popular posts criticising extremist parties or damage the FreeWill politicians by injecting fake posts under their name.

6.2.2 Reflection in MISP

This section shows how different organisations report operational CTI regarding the story of election interference described into the SPARTA MISP instance.

Organisations CESNET, LKA, LEO, NASK, INOV, KTU and EUT are all connected to the sharing community (MISP instance). Each organisation uses its unique methods to detect cyber incidents in its field and if it figures out something worth sharing, it shares it. There is also a (fictional) organisation CERT.RPE (national CERT of the Republic of Peripheral Europe) which focuses on a more global view. It observes the shared events, tries to find correlations and spot potential threats to national security (including election interference).

CESNET detects backscatter from a DDoS attack against IP address 12.12.13.13. The hostname associated with this IP address is “www.freewillparty.rpe”. CESNET creates a corresponding event in the sharing platform. The process of event creation, as well as the final event, is shown in the screenshots below. The event contains the targeted IP address and port number, domain name, the approximate volume of the attack and time information. Due to the detection mechanism, no information about attack sources is known. To better mark the event, DDoS-related tags are added. Since the attack is obviously election-related (web pages of a political party is attacked), the event is assigned to a corresponding MISP galaxy.

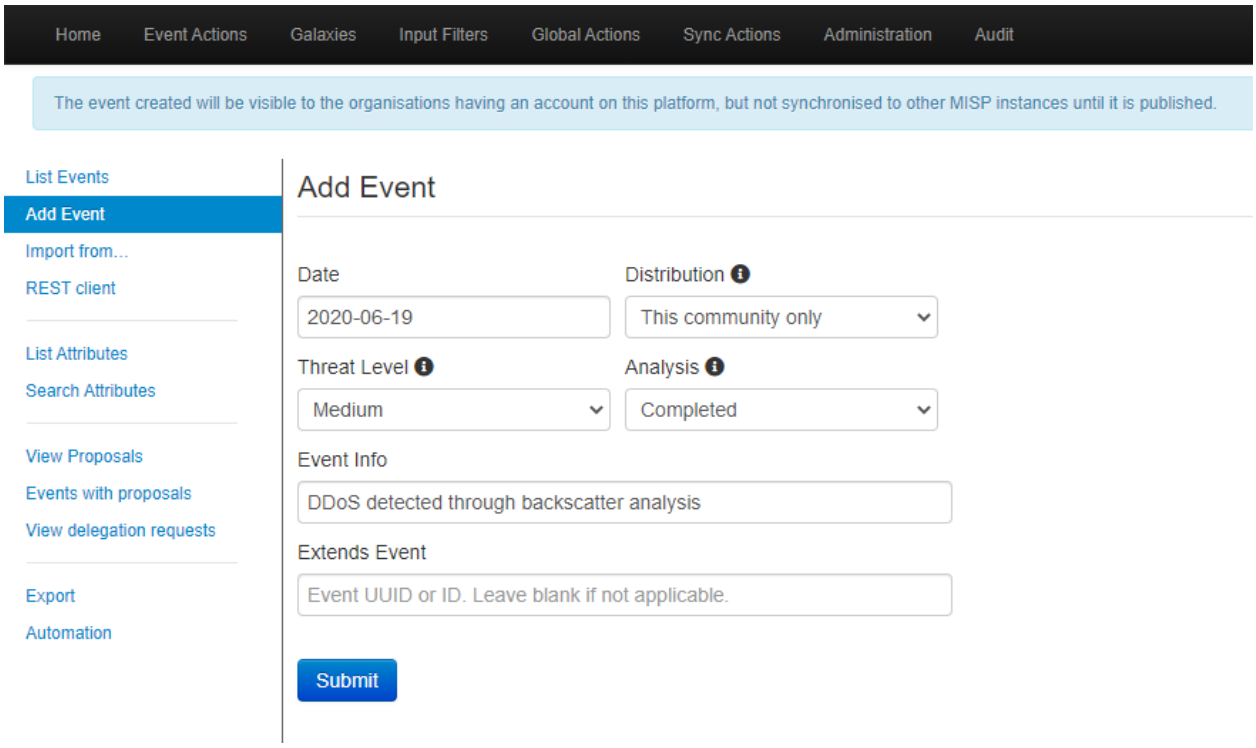

Figure 6.1: DDoS event step 1

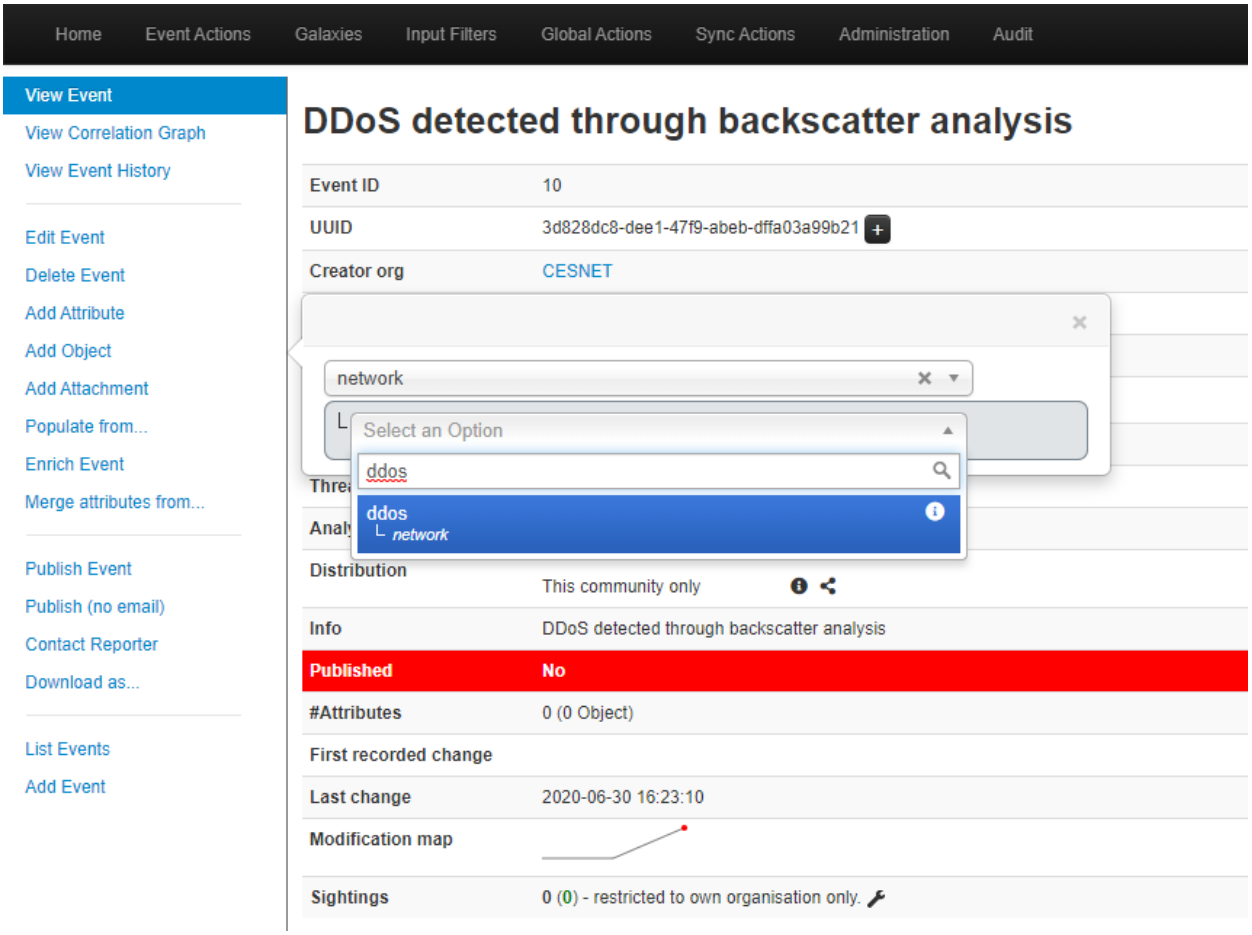

Figure 6.2: DDoS event step 2

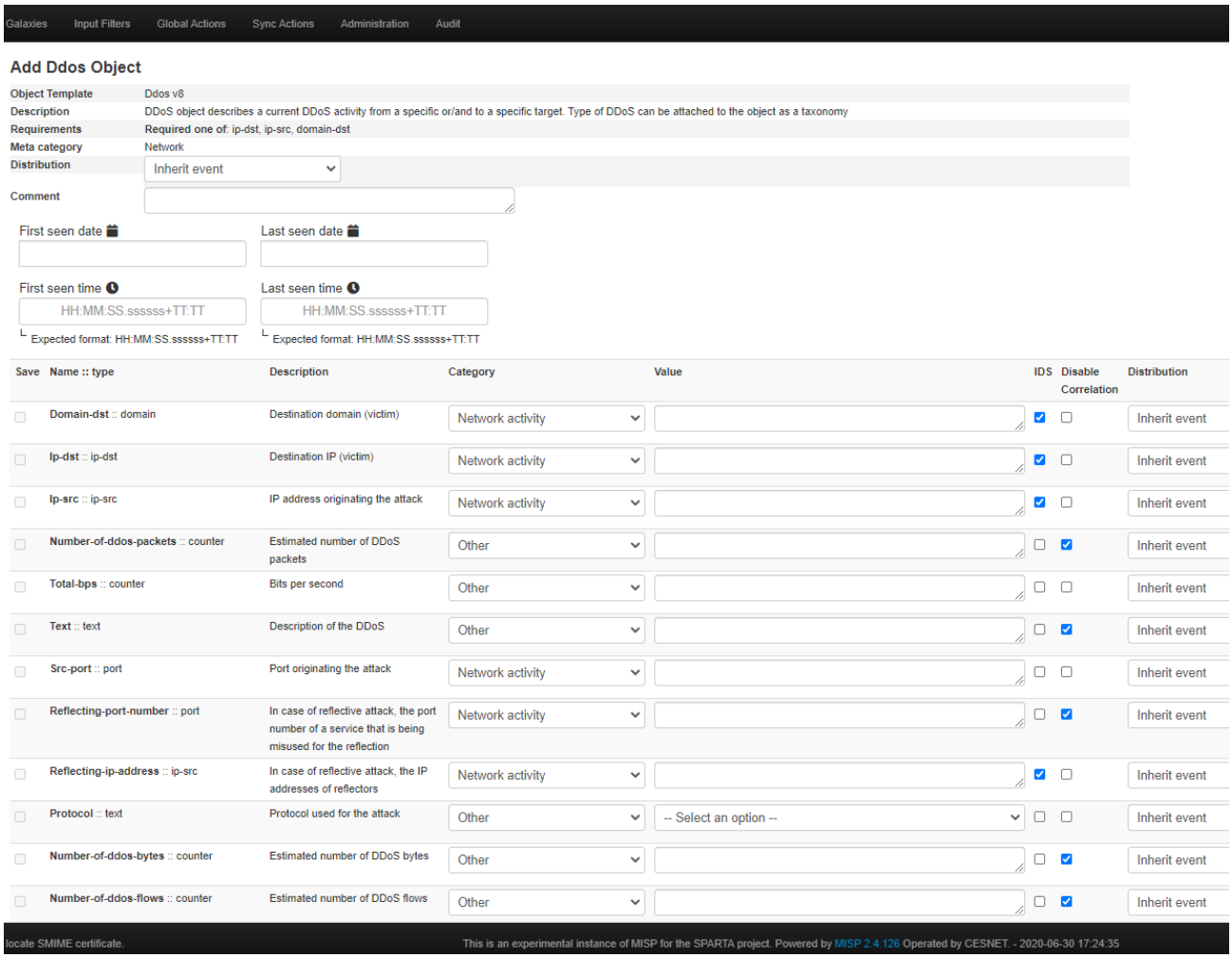

Figure 6.3: DDoS event step 3

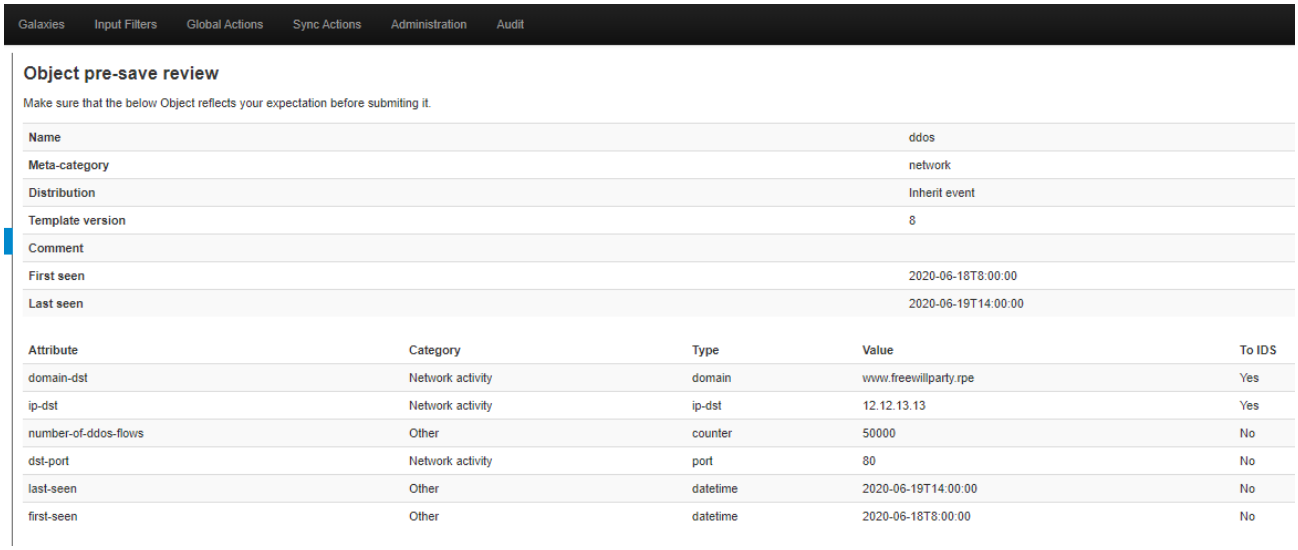

Figure 6.4: DDoS event step 4

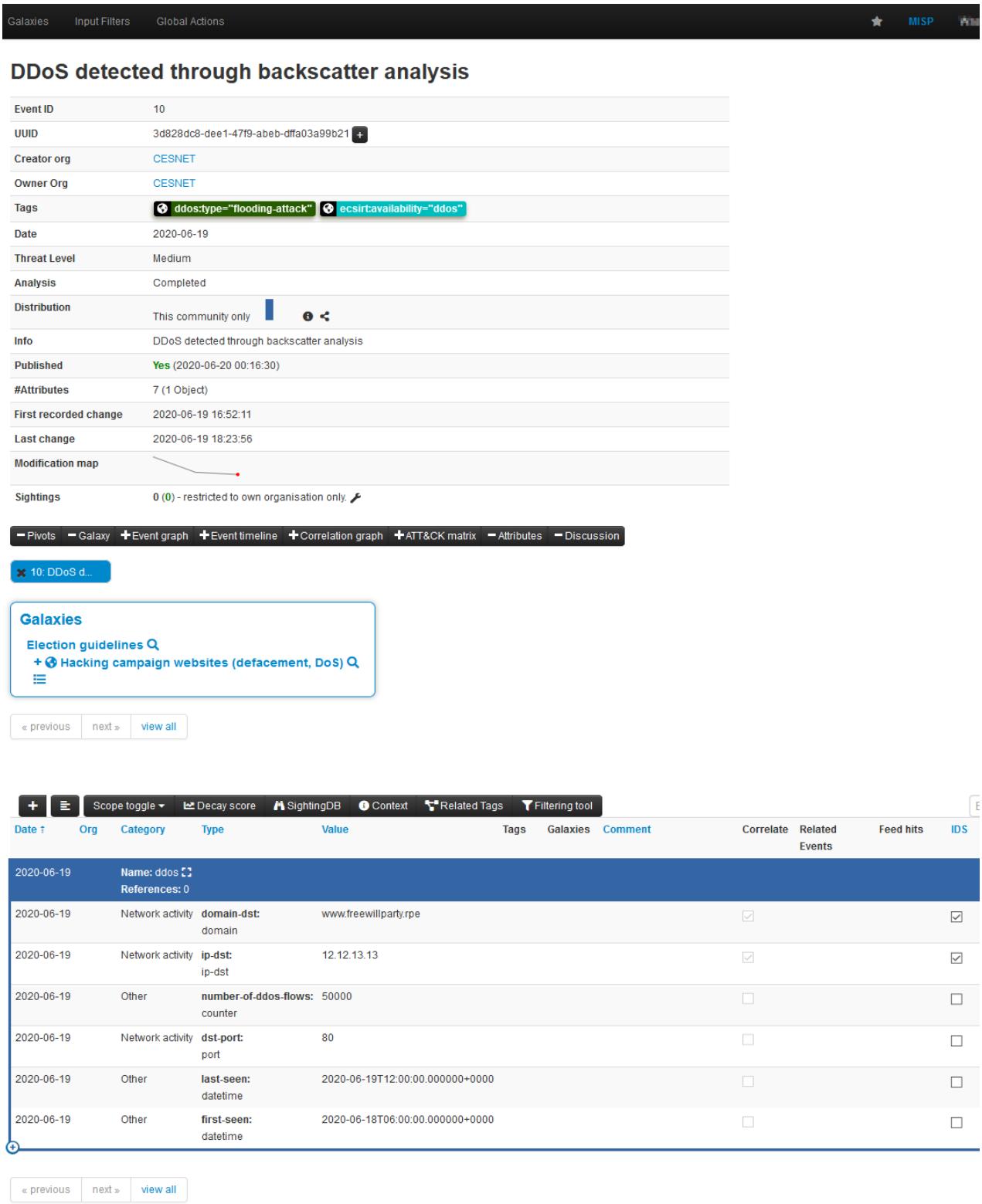

Figure 6.5: Full DDoS event

Shortly, KTU notices the information about a DDoS attack in MISP and tries to find the corresponding traffic in their flow data. Fortunately, a part of the distributed attack goes through their monitored network infrastructure, so they are able to extract more details about the characteristics of the attack and several source IP addresses. KTU proposes new attributes to CESNET's event describing these new findings. CESNET accepts the changes. The creation of the proposal and the event with proposed changes are shown in the screenshots below.

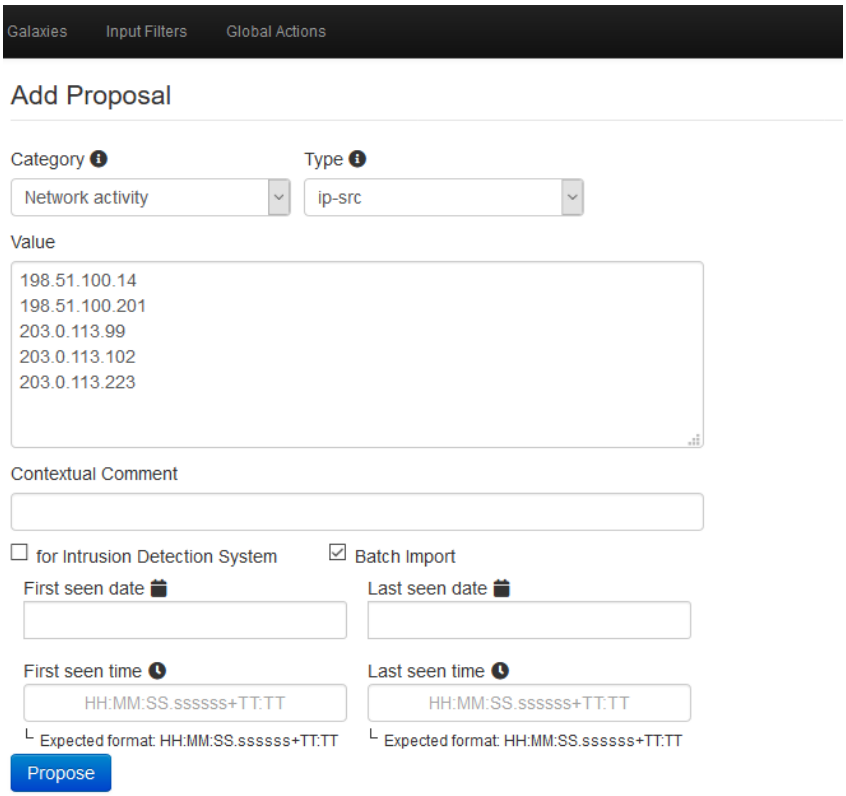

Figure 6.6: DDoS event proposal step 1

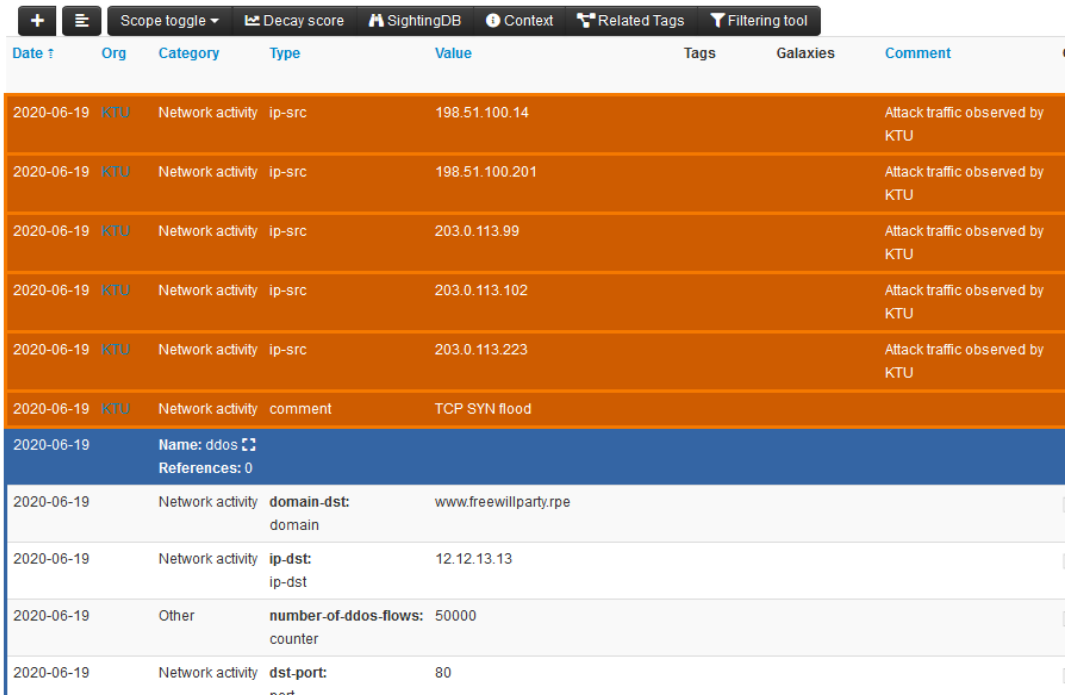

Figure 6.7: DDoS event proposal step 2

In the meantime, LEO notices a spam campaign containing suspicious links to the fake URLs of web pages that look similar to the original FreeWill party's web pages. LEO pushes their findings as an event to the sharing platform and marks it by an appropriate tag and galaxy. Details of the event are shown on the screenshot below.

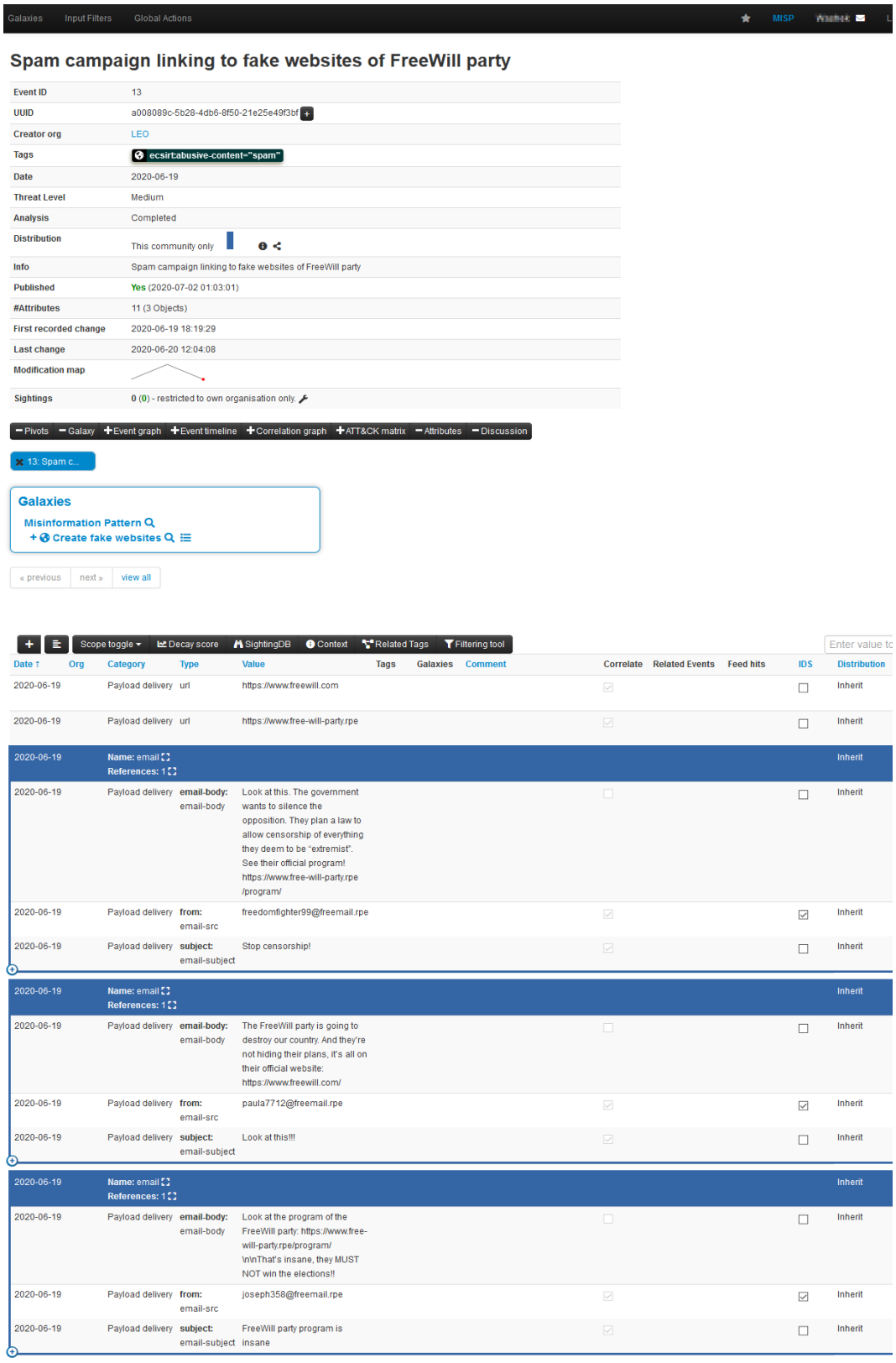

Figure 6.8: Spam event

CNR detects the same fake links to be posted and promoted by several Twitter accounts known to be extremists' trolls or fake accounts. Therefore, CNR creates an event in the sharing platform and includes the fake web URLs, IDs of the twitter accounts, and examples of the tweets. It also assigns the corresponding galaxies. The event is shown in the screenshot below.

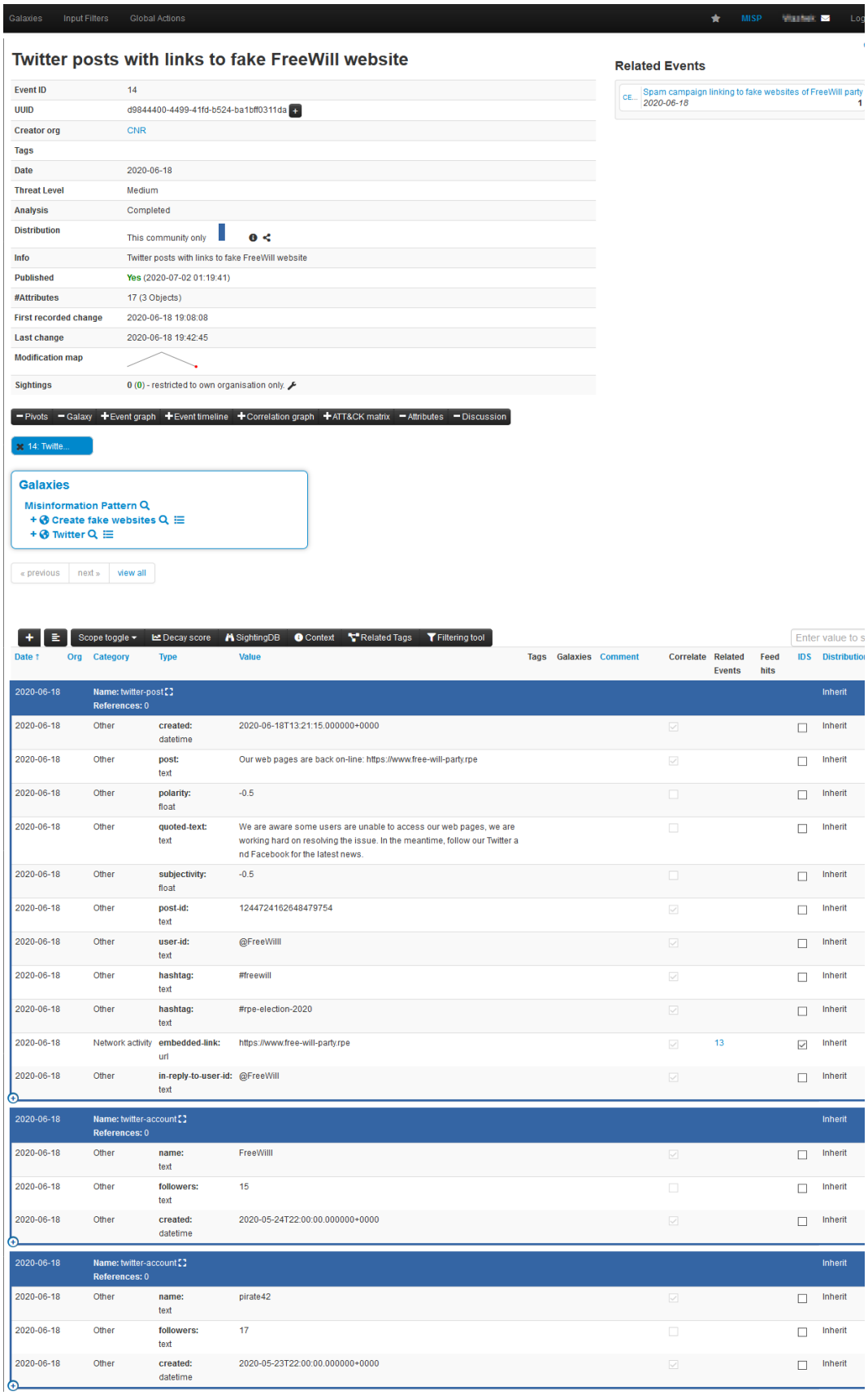

Figure 6.9: Twitter event

As can be seen in the Figure 6.10: Spam event related with the Twitter event image, MISP automatically marks the two events as correlated since they contain the same attributes - links to the fake website.

Figure 6.10: Spam event related with the Twitter event

Later on, CERT.RPE notices all these events and deduces that they are all part of an election interference campaign. They create a new event in MISP whose purpose is to link together all events related to the RPE parliament election. The event is named “Interference of parliament elections in the Republic of Peripheral Europe” and a single object of type “election-interference-campaign” is added to its attributes (see the screenshot below).

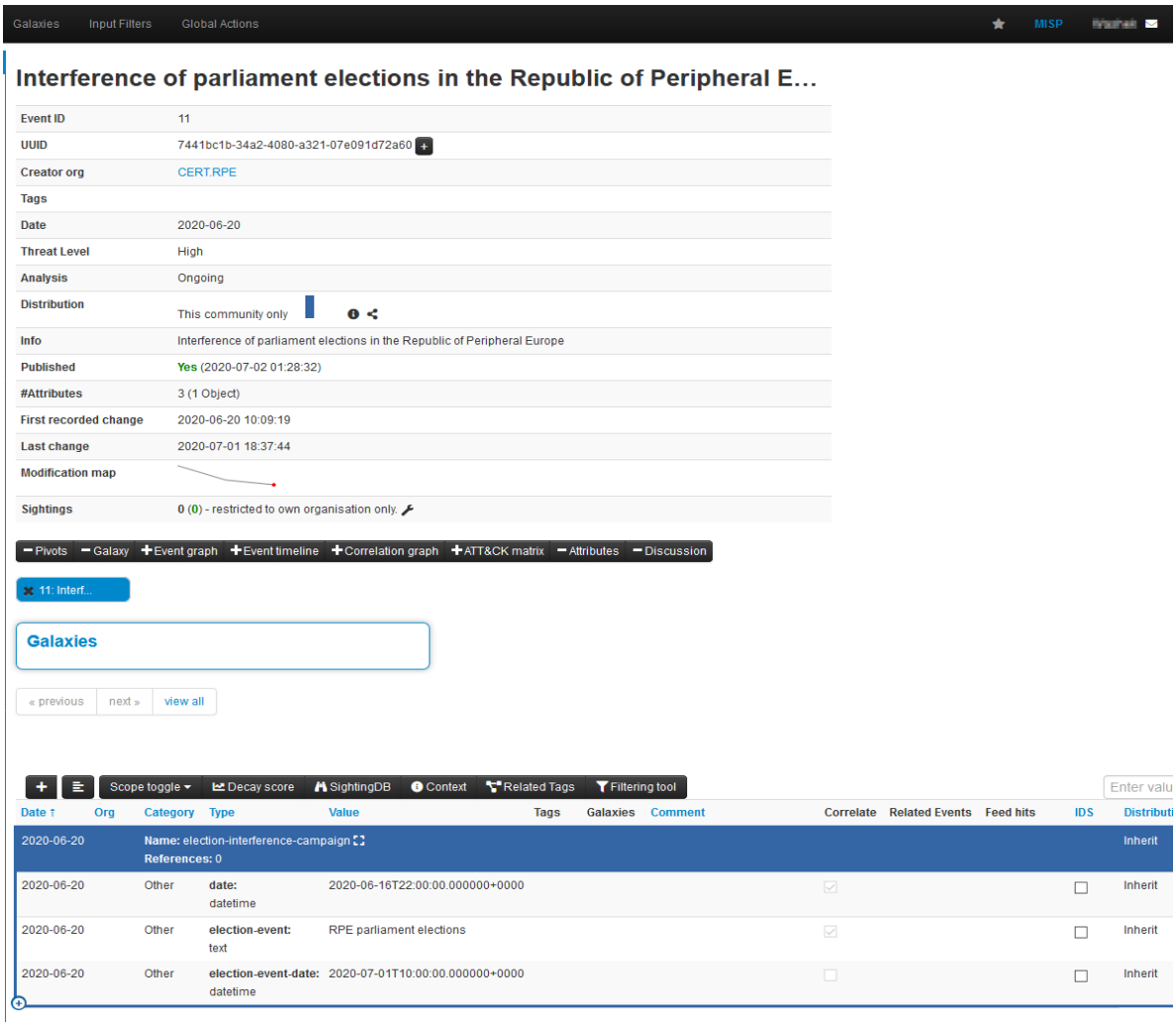

Figure 6.11: Election interference event

The “Extends event” functionality of MISP should be now used to link the DDoS, spam and twitter events to this global event. However, this parameter can only be set by the authors of the

individual events, so CERT.RPE contacts them (using a MISP built-in functionality “Contact author”) and they eventually add the relationship.

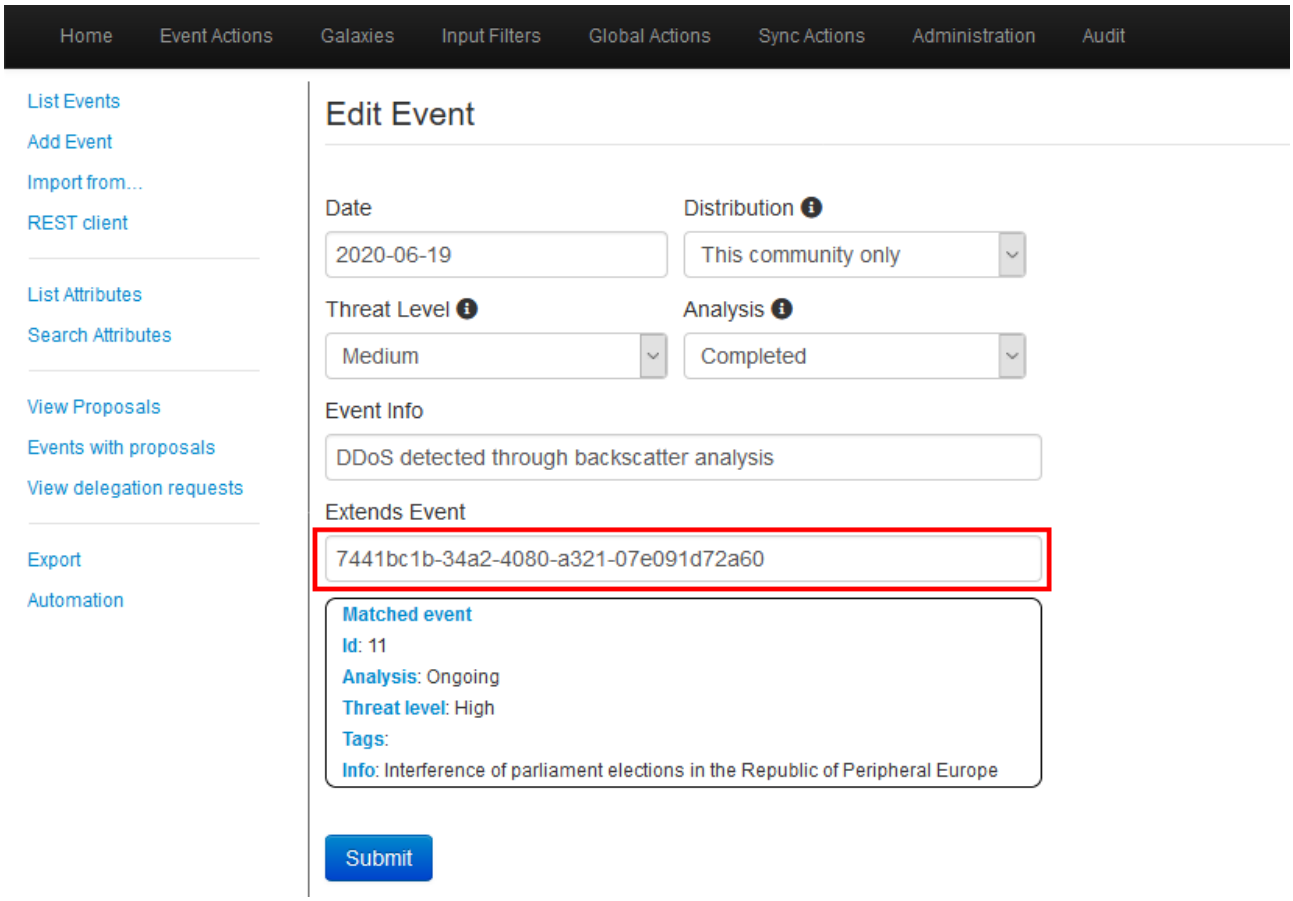

Figure 6.12: Editing DDoS event to extend election interference event

Figure 6.13: Additional events linked with election interference event

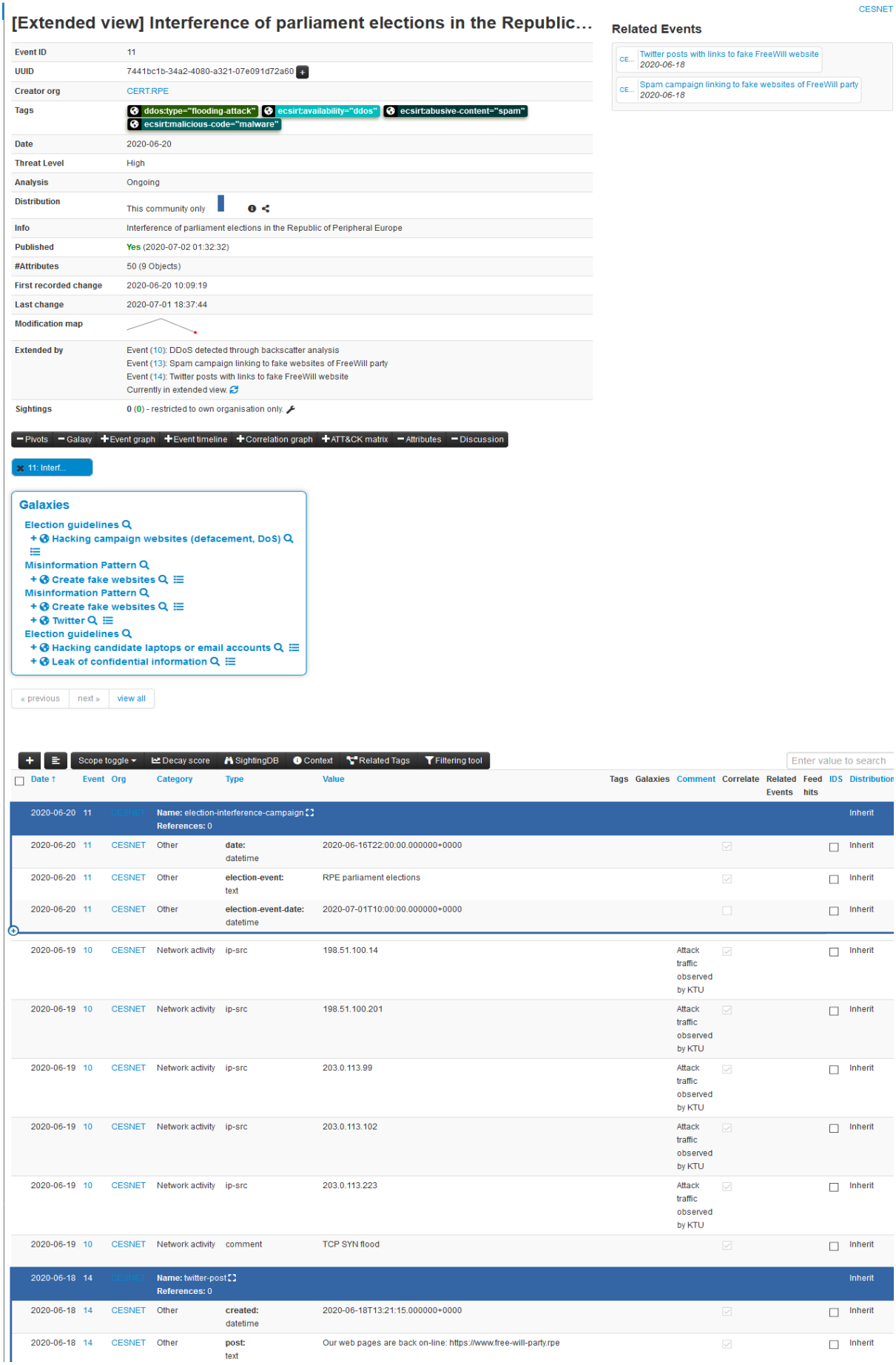

Figure 6.14: Extended view of event with extending events

MISP then allows switching to the “extended view” in which a user can see all the tags, galaxies, attributes and objects from the linked events within the main/global event.

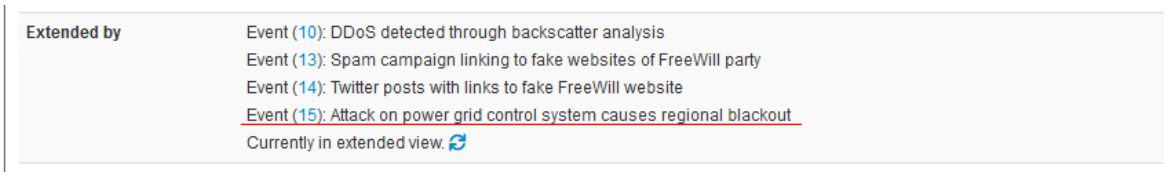

Figure 6.15: List of events extending election interference event

INOV operates intrusion detection systems in several critical infrastructures in the country. They detect anomalies in one of the power grid control systems. Initial analysis of the anomalies shows that it might cause some problems with power supply. Still, it is not deemed critical (note: attackers are just testing access to different systems; it is not the main attack yet). Nevertheless, the IDS alerts are shared to the MISP instance via an event labelled with the “Energy” sector galaxy. The corresponding event is shown below.

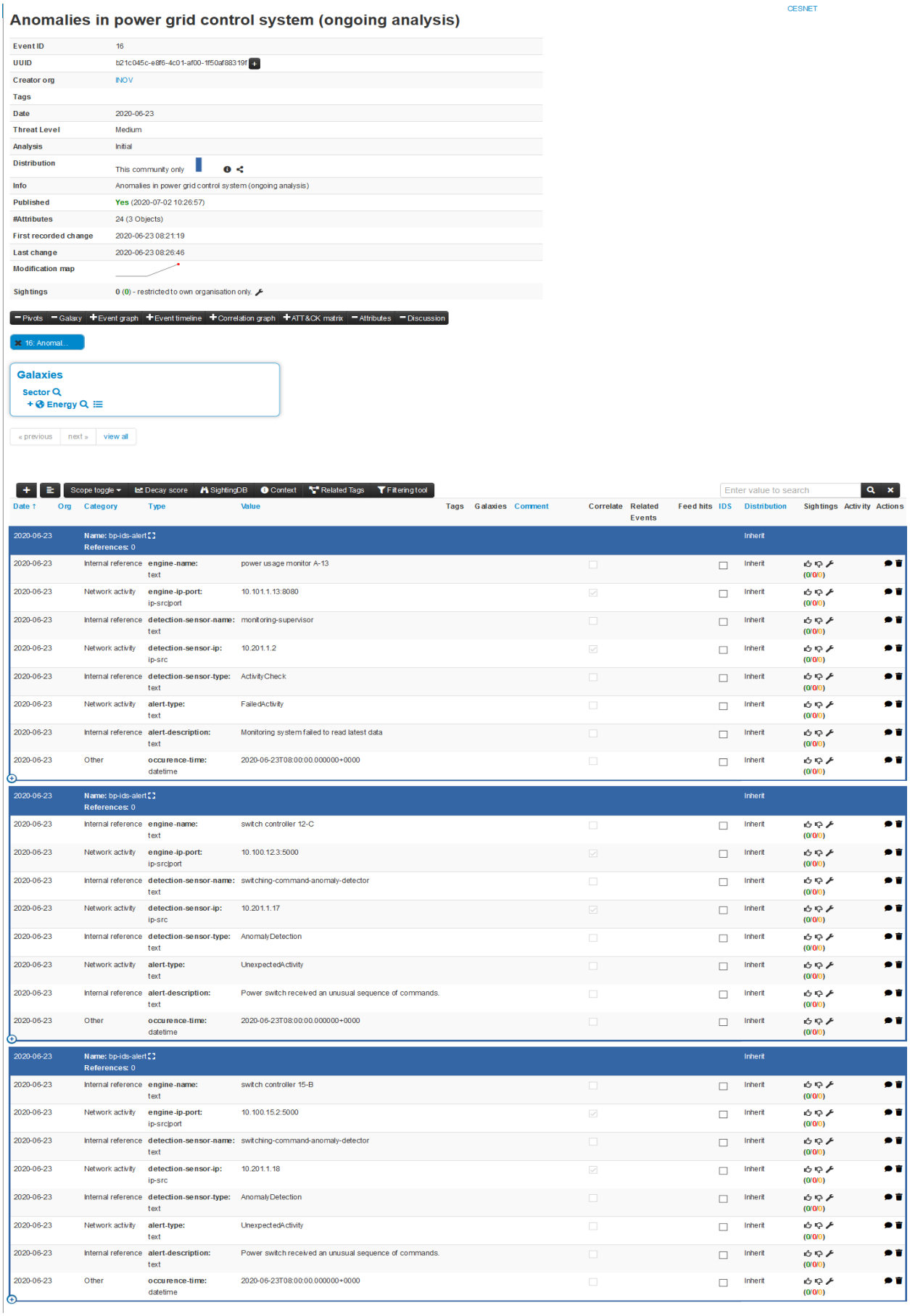

Figure 6.16: BP-IDS alert event

All events shared into the MISP instance are automatically analysed by EUT's AI system trying to find correlations, either between the MISP events themselves or with some other information sources. One of the data sources used is a list of upcoming real-world events that could be cyber-attacked, as compiled by the web-scraping system of LKA. The AI system notices a correlation between the power grid anomaly and the ceremonial opening of the new solar power plant which should happen in just a few hours (and which is included in the LKA's list). Therefore, it proposes a new tag to be added to the event - current-event:energy="solar-park-opening". Note: It is currently not possible to directly propose tags in MISP, although there is a request for such functionality on MISP's GitHub, so we propose an attribute of "Other/text" type with the tag as its content, instead.

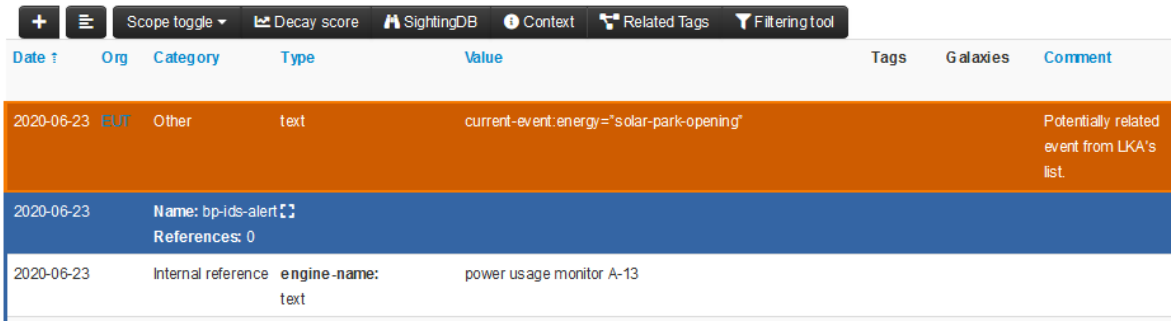

Figure 6.17: BP-IDS event enriched

In the meantime, more detailed analysis of the anomalies by INOV suggests the problem is more severe than initially anticipated. The correlation found by EUT further helps analysts to realize what is going on and they immediately warn power grid operators about the imminent threat. The attack is triggered shortly after it and although the operators are not able to prevent the blackout, at least the warning helps them to limit its impact to a small region. INOV updates the description of the event and its severity.

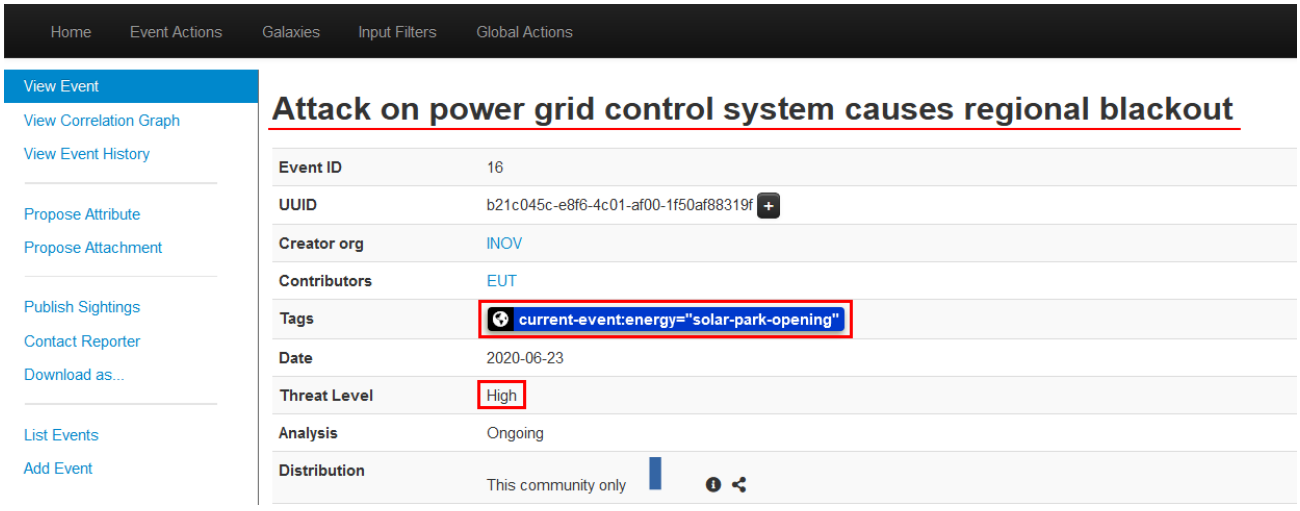

Figure 6.18: BP-IDS alert updated

FreeWill party has a suspicion about leaks of sensitive data from their internal IT systems, so they order a security audit from CyberSecGurus company. They find an unknown malware residing on one of the servers and give it to NASK for analysis.
NASK finds out that the malware specifically targets the FreeWill systems and its main purpose is to capture credentials and exfiltrate sensitive data. The malware sample is new, but NASK finds

out it is derived from the “Agent Edison” spyware. NASK also finds some similarities with a few other samples they analysed before, indicating they may come from the same threat actor (although it is unknown to NASK which actor it is). With permission from CyberSecGurus and FreeWill party, all these information are shared to MISP. The corresponding event is shown below. A complete binary of the sample is included as well. Since the malware targets a political party, appropriate tags and galaxies are added and the event is linked to the “Election interference” event of CERT.RPE.

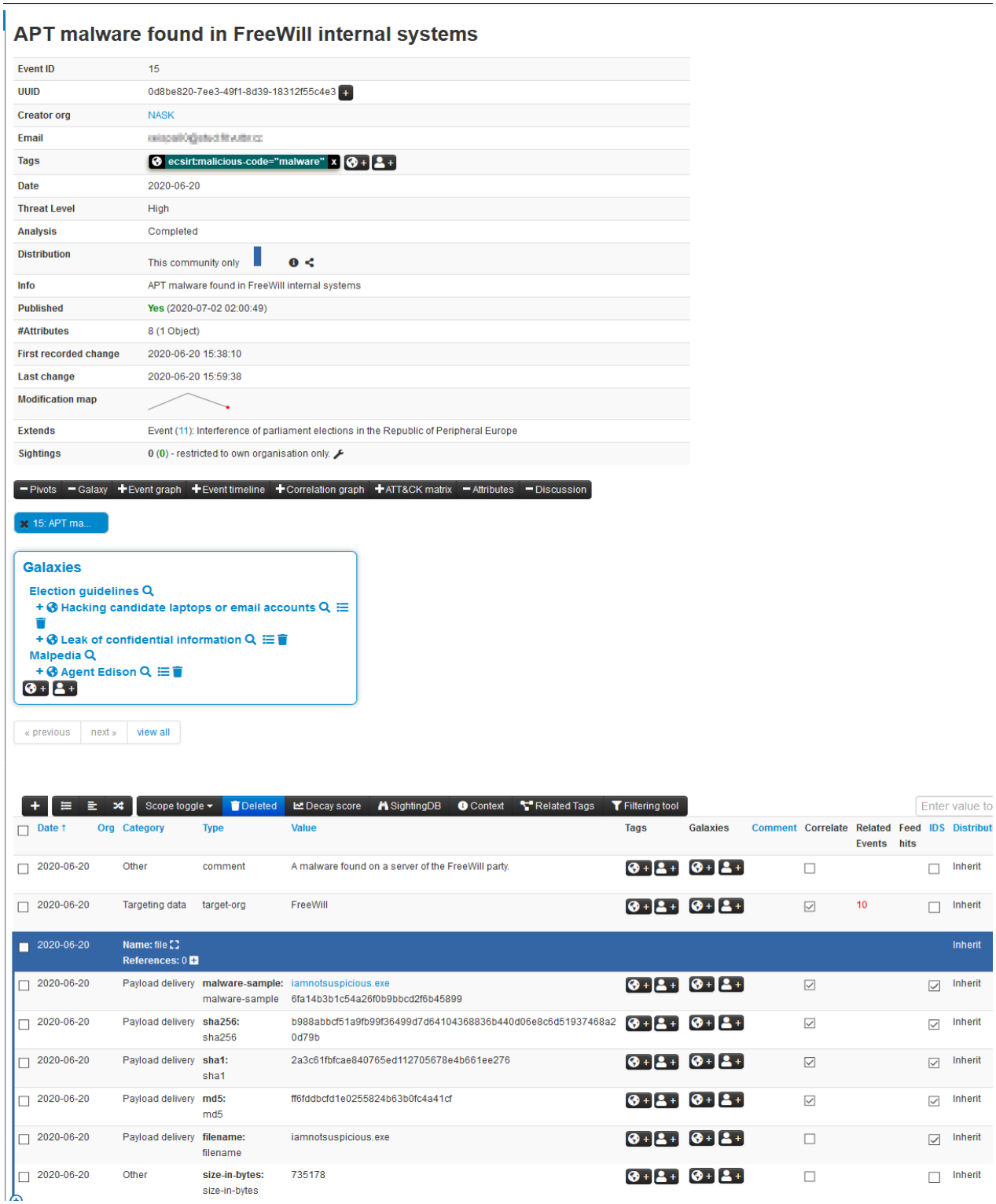

Figure 6.19: Malware event of the first sample

EUT correlates the information about the malware with its own intelligence and finds out that the malware HackAllWorlds group probably created the malware. Therefore, EUT proposes to add the following attribute to the MISP event. NASK trusts EUT and accepts the proposal.

Figure 6.20: Malware event enriched

NASK also analyses another malware sample, which was found on machines of several users in the Republic of Peripheral Europe. It was found out that it is able to manipulate the content of web ads served by major advertising networks and that it focuses on political and election-related ads. They also found artefacts suggesting it is able to do similar manipulations with social networks posts. The result of the analysis was shared into MISP as the event shown below. Since the malware manipulates with political ads, the event was linked to the “Election interference” event of CERT.RPE.

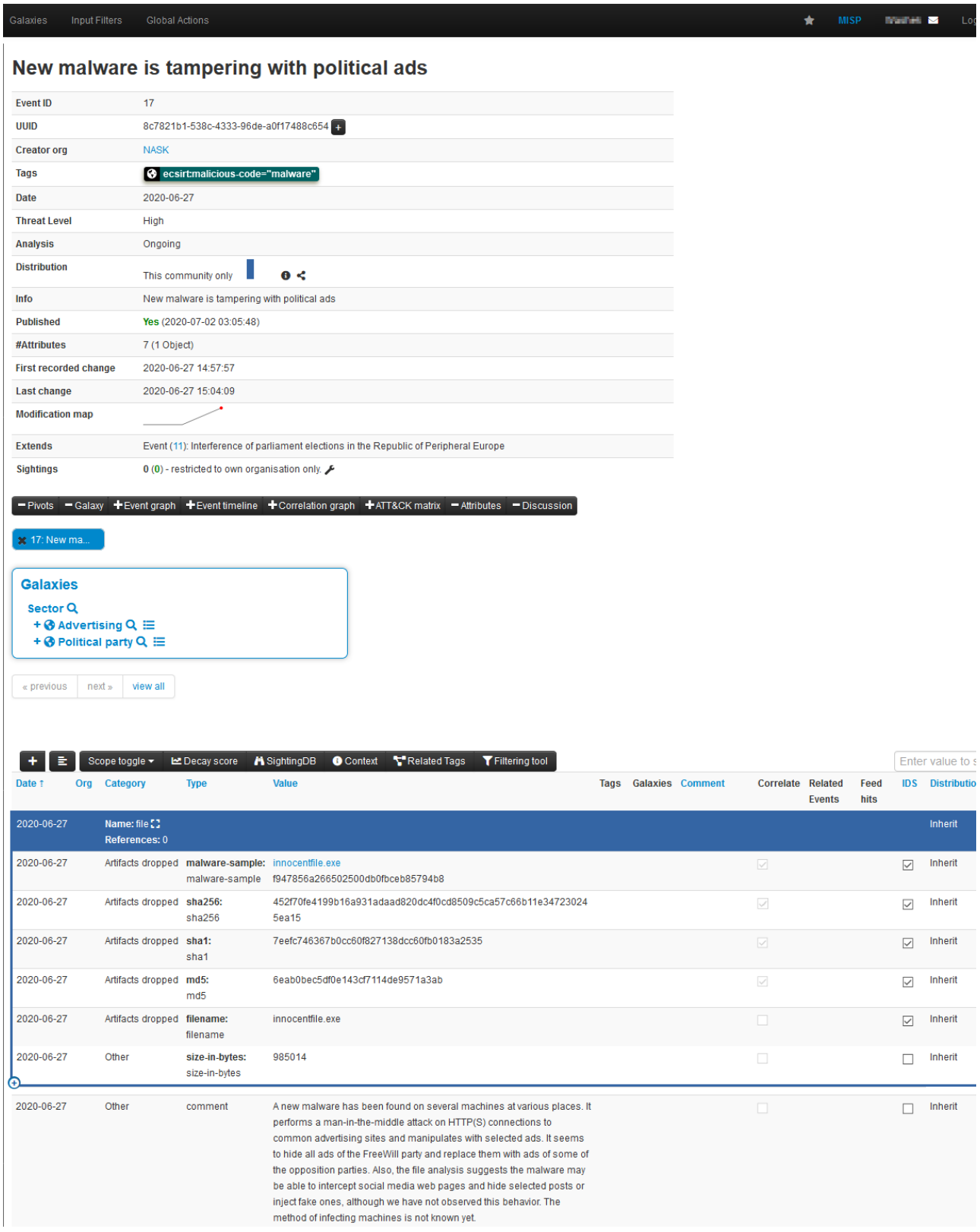

Figure 6.21: Malware event of the second sample

To summarise, the MISP instance contains multiple events related to the described election interference scenario, each reporting a different part of the story. A high-level event created by the national CERT (CERT.RPE) links the others together. Although some of the Radical Pirates' actions remained hidden, it is clear that there was a complex campaign to interfere with elections.

Traces found in one of the malware samples lead to the HackAllWorlds group. It is now on law enforcement agencies to find the connection to Radical Pirates.

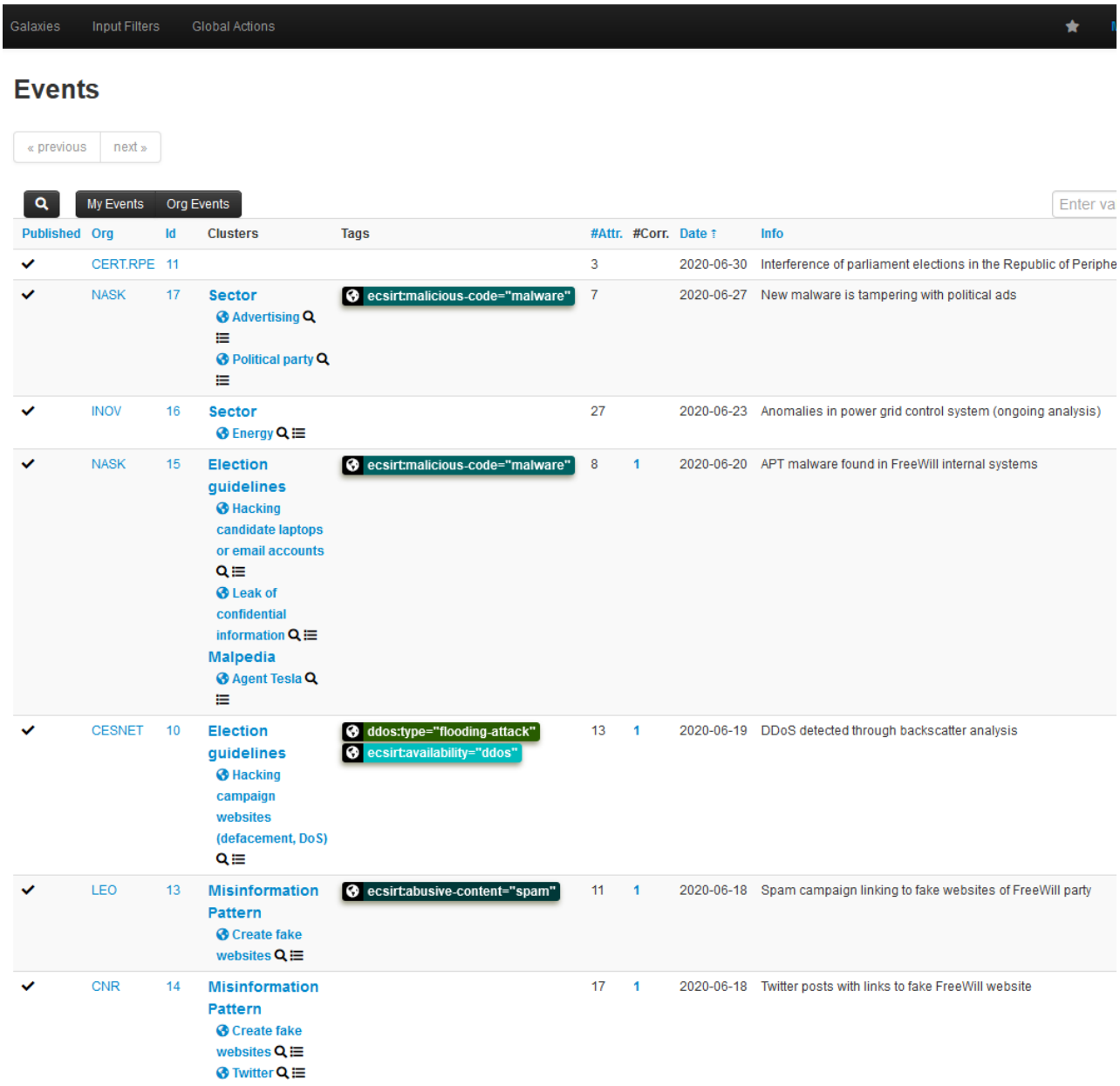

Figure 6.22: List of all events

Chapter 7 Plans and roadmap

This section provides a structured overview of planned upcoming activities within the Project in the utilisation of treat intelligence common data model and beyond the Project.

During the first year of T-SHARK implementation, several activities have been taking place concerning the cybersecurity threat intelligence common data model. As one of them – analysis of threat intelligence data exchange platforms for sharing, storing and correlating. Although we choose MISP as a common data model, we further review the platforms, to see what functional and structural concepts they are missing and can be extend with. The evaluation is on-going with other tools and further alignment and integration activities should be foreseen during the second half of T-SHARK programme implementation (1).

> *MISP, as already mentioned can be used as threat intelligence platform for sharing, storing and correlating Indicators of Compromise of targeted attacks, threat intelligence, financial fraud information, vulnerability information or even counter-terrorism information, training and education. MISP built-in sharing functionality is used to ease data sharing using a different model of distributions and it can synchronise events and attributes automatically among different MISP nodes. Advanced filtering functionalities can be used to meet each organisation sharing policy, including a flexible sharing group capacity and an attribute level distribution mechanism.*
>
> *In MISP you can bulk-import, batch-import, free-text import, import from OpenIOC, GFI sandbox, ThreatConnect CSV or MISP format from various data sources.*
>
> *It is also a flexible tool to import and integrate MISP feed and any threat intel or OSINT feed from third parties. Of course, many default feeds are included in standard MISP installation and it can be used as a basis to extend guidelines and rules of the Comprehensive CTI model into materials, focused on training and education.*

Sharing of raw data, as well as intelligence information and insights, is a strategic resource of successful cybersecurity operations. Since attackers share information cross the boarders and among their peers too, therefore it is essential for cybersecurity community to share, exchange, integrate information in order to counter these complex and heterogeneous cases successfully and stay informed on new emerging threats. A comprehensive threat intelligence sharing platform, where users from the cybersecurity community and other communities at large, can share their information on incidents or other artefacts in an efficient, trusted environment can be enabled by common cybersecurity threat intelligence data model. The integrated complex scenario is the best to validate it. In T-SHARK it is planned for final Stage #3 presentation where common data model will be used and processed in the context of Umbrella Demo Case (2).

Future work and plan of development beyond SPARTA are manifold. In a future iteration process, the cross-platform replication and synchronisation protocol will be analysed for its efficiency, gaps and integration possibilities. Another step is the quality improvement of the shared information and potential analytical extension, respectively, information structures and embedded taxonomies are strategic next step for improvement (3).

Next possible quality improvement is adoption for the processing of the large datasets that are generated by various stakeholders of comprehensive cybersecurity incident processing. Big data mining techniques are essential for predictive analysis; therefore, data model extension to support that is natural next step of evolution (4).

Also, visual analytics techniques is important for supporting of strategic decisions, high-level situational awareness cases, organisation of table top exercises, gamification part of comprehensive cybersecurity incidents analysis. All these aspects can be intergraded into Cyber ranges. Adoption and integration into major cyber ranges platforms have its own huge potential and exploitation internally within SPARTA community and on a global scale. That can be a brave new direction for T-SHARK Common Data Model (5).

The core of the common data model is built from practical use-cases in information security, intelligence communities, incident response teams and strategic analysis stakeholders' groups. These groups provide different perspectives on practical applicability and niche optimisation while processing different analysis methods. As the various sub-cases are going to be developed further and integrated until the final DEMO of T-SHARK, the common data model will naturally evolve and extend in alignment with subcases. It will maintain the core part, identify required new features and prioritises the implementation based on the impact and usefulness in the different scenarios of the application represented by subcases (6).

Further development and utilisation of the common data model are hardly imaginable without the emergence of main governance and management principles (7):

- data format / standard governance principles
- adoptive maintenance responding to changing environment needs
- introduction and onboarding for new users
- versioning and compatibility principles

Knowledge is another key driver for success and efficient use of the data model. Creating competence development requirements and integrating them into training, extending users guidelines and rules of the comprehensive CTI data model into materials of SPARTA training activities as well as standard MISP knowledge libraries. (8)

The list of upcoming activities is presented further. It provides the general roadmap, including activities within the Project, also covers activity beyond it. It is to be noted, that it might be slightly changed, as activities are highly dependent on developments within Sub-cases and other EU projects that are intended to be included in Stage Gates process and integrated to comprehensive cybersecurity concept.

Activities, planned for the future cybersecurity threat intelligence common data model roadmap implementation:

#	Activity/Planned item	Target period
(1)	Evaluation of identified sharing platforms	SPARTA M18
(2)	Adoption for final Umbrella Demo Case	SPARTA M36
(3)	Maturation of the identified sharing technologies	SPARTA M23
(4)	Large data sets and big data predictive analytic extension	Beyond SPARTA
(5)	Extensions for Cyber Ranges platforms	Beyond SPARTA
(6)	Integration of the selected sharing technologies	SPARTA M30, M36
(7)	Main governance and management principles	SPARTA M36
(8)	Data model trainings and integration	SPARTA M36
(9)	Presentation to industry, academia, end-users ecosystem	Beyond SPARTA
(10)	Propagation for international and national experts communities	Beyond SPARTA

(11)	Ecosystem establishment	Beyond SPARTA
(12)	Inter-pilot cooperation	Beyond SPARTA
(13)	Ontology for mapping of domains	Beyond SPARTA

Within the SPARTA project scope:

- evolution in alignment with subcases
- adoption Analysis Framework
- adoption to Visual analytics
- integration into cross information exchange platform 4.4
- evaluation of C3ISP, MISP, OTX, Protective platforms
- integration into all data source intelligence
- creating competence through training, extending guidelines and rules of the comprehensive CTI model into materials for MISP as transferring them through WP focused on training and education

Activities, related to partners of the Project:

- validation of individual solutions data exchange
- validation of individual solutions internal processes

Activities, foreseen beyond the Project:

- presentation to industry, academia, end-users ecosystem
- propagation for international and national experts communities
- ecosystem
- inter-pilot cooperation
- further extension of the data sources domains
- further extension into not only operational CTI
- ontology for mapping of various information domains

Chapter 8 Summary and Conclusion

This deliverable documents our approach to address the challenge of the CTI common data model. We took a practical approach and we focused on a particular field of CTI, namely, operational CTI where we benefited from the on-going activities within the T-SHARK Programme. We derived requirements from the developed use cases related to election interference. Together with the collected related work, this served as an input for our strategic choice. The outcome of the strategic choice resulted into the proposal of a methodological approach to build data models of use cases consistently to support machine readability as well as to the need for ontology to enable the mapping of models. We demonstrated the applicability of the proposed methodology on practical use cases as well as by deploying SPARTA MISP instance and populating it with the extended models immediately. During our investigation, we encountered various topics and areas for further research and development and these created our future plan and roadmap within T-SHARK Programme but also beyond.

Chapter 9 Bibliography

[1] Farnham, G., Leune, K., 2013. Tools and Standards for Cyber Threat Intelligence Projects.
[2] Sauerwein, C., Sillaber, C., Mussmann, A., Breu, R, 2017. Threat intelligence sharing platforms: an exploratory study of software vendors and research perspectives. Towards Thought Leadership in Digital Transformation: 13. Internationale Tagung Wirtschaftsinformatik, WI 2017, St.Gallen, Switzerland, February 12–15, 2017.
[3] Gong N., 2019. Barriers to Adopting Interoperability Standards for Cyber Threat Intelligence Sharing: An Exploratory Study. In: Arai K., Kapoor S., Bhatia R. (eds) Intelligent Computing. SAI 2018. Advances in Intelligent Systems and Computing, vol. 857. Springer, Cham.
[4] Fransen, Frank, Andre Smulders, and Richard Kerkdijk. "Cyber security information exchange to gain insight into the effects of cyber threats and incidents." e & i Elektrotechnik und Informationstechnik 132, no. 2 (2015): 106-112.
[5] Ussath, Martin, David Jaeger, Feng Cheng, and Christoph Meinel. "Pushing the limits of cyber threat intelligence: extending STIX to support complex patterns." In Information Technology: New Generations, pp. 213-225. Springer, Cham, 2016.
[6] Martinelli, Fabio, Oleksii Osliak, and Andrea Saracino. "Towards general scheme for data sharing agreements empowering privacy-preserving data analysis of structured CTI." In Computer Security, pp. 192-212. Springer, Cham, 2018.
[7] Osliak, Oleksii, Andrea Saracino, and Fabio Martinelli. "A scheme for the sticky policy representation supporting secure cyber-threat intelligence analysis and sharing." Information & Computer Security (2019).
[8] Institute, P., 2017. Third Annual Study on Exchanging Cyber Threat Intelligence: There Has to Be a Better Way. https://www.ponemon.org/local/upload/file/2017%20Inflobox%20Report%20V6.pdf
[9] Pawlinski, P., Jaroszewski, P., Kijewski, P., Siewierski, L., Jacewicz, P., Zielony, P., Zuber, R., 2014. Actionable information for security incident response. European Union Agency for Network and Information Security, Heraklion, Greece.
[10] Brown, S., Gommers, J., Serrano, O., 2015. From cyber security information sharing to threat management. In: Proceedings of the 2nd ACM Workshop on Information Sharing and Collaborative Security. ACM, pp. 43–49.
[11] Al-Ibrahim, O., Mohaisen, A., Kamhoua, C., Kwiat, K., Njilla, L., 2017. Beyond free riding: quality of indicators for assessing participation in information sharing for threat intelligence. arXiv: 1702.00552.
[12] Mohaisen, A., Al-Ibrahim, O., Kamhoua, C., Kwiat, K., Njilla, L., 2017. Rethinking information sharing for actionable threat intelligence. CoRR. arXiv: 1702.00548.
[13] Alliance, H.I.T., 2015. Health Industry Cyber Threat Information Sharing and Analysis. Technical Report.
[14] ThreatConnect, 2015. Threat Intelligence Platforms - Everything You've Ever Wanted to Know But Didn't Know to Ask. Technical Report.
[15] Sweeney, L., 2002. k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl. Based Syst. 10 (5), 557–570.
[16] Machanavajjhala, A., Kifer, D., Gehrke, J., Venkitasubramaniam, M., 2007. L -diversity: privacy beyond k -anonymity. TKDD 1 (1), 3.
[17] Li, N., Li, T., Venkatasubramanian, S., 2007. t-closeness: privacy beyond k-anonymity and l-diversity. In: Proceedings of the 23rd International Conference on Data Engineering, ICDE 2007, The Marmara Hotel, Istanbul, Turkey, April 15–20, 2007, pp. 106–115.

[18] Garrido-Pelaz, R., González-Manzano, L., Pastrana, S., 2016. Shall we collaborate? A model to analyse the benefits of information sharing. In: Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security. ACM, pp. 15–24.

[19] de Fuentes, J.M., González-Manzano, L., Tapiador, J., Peris-Lopez, P.,2016. PRACIS: privacy-preserving and aggregatable cybersecurity information sharing. Comput. Secur.

[20] Best, D.M., Bhatia, J., Peterson, E.S., Breaux, T.D., 2017. Improved cyber threat indicator sharing by scoring privacy risk. In: Technologies for Homeland Security (HST), 2017 IEEE International Symposium on. IEEE, pp. 1–5.

[21] P. Nespoli, D. Papamartzivanos, F. Gómez Mármol and G. Kambourakis, "Optimal Countermeasures Selection Against Cyber Attacks: A Comprehensive Survey on Reaction Frameworks," in IEEE Communications Surveys & Tutorials, vol. 20, no. 2, pp. 1361-1396, Secondquarter 2018, doi: 10.1109/COMST.2017.2781126.

[22] Sammut-Bonnici, Tanya & Galea, David. (2015). SWOT Analysis. 10.1002/9781118785317.weom120103

Chapter 10 Annex – A: Requirements

All the partners of T-Shark were asked to provide requirements from their perspective and there were two additional iterations for review, editing and adding of new requirements.

Table 10.1: Requirements

Author	*Status*	ID	Name	Priority	Description	Evaluation
LIST+EUT	*OK*	DCM-GEN-EXT-01	Extensibility	MUST	The data model must allow including extensions to the standard specification. The extension is considered to be an additional value in case of an enumerated list of values or additional data structure. The extensions will be used by vendors or communities to represent specific data or relations unique to their proprietary use cases, or will allow the model to evolve according to the future needs, i.e. the extensions will serve as a proof-of-concept and will become standard in future versions of the data model. The extended data model must be backwards-compatible with the standard specification, i.e. a third party which is not aware of the specification of the extension must be able to work with the data model including the extended part but is not capable of understanding the semantic of the extended part.	The data model allows using unknown values without affecting the capability of understanding the semantic of the standard part of the data model. The data model allows to define new data structures and link these data structures into the existing data model without affecting the capability of understanding the semantic of the standard part of the data model

Author	Status	ID	Name	Priority	Description	Evaluation
LIST+EUT	*OK*	DCM--STD-01	Standardisation	MUST	The data model must be adjusted to different existing standards to model threat and attack intelligence that is currently in use, or being deployed. This requirement is necessary in order to be able to gather information from different sources using these sorts of standards. In addition, standardisation of the data model would give it a special provision for the widest possible adoption.	It is feasible to validate the different standards adopted against the schemes defined for each of them.
LIST+EUT	*OK*	DCM-GEN-VAL-01	Check for validity	MUST	The data model must be capable of being (automatically) checked against grammatical and construction rules, and may be declared as valid or non-valid. This property will ensure that data is exchanged according to an expected schema that can be understood by all, and will facilitate the sharing of information.	It is possible to launch a validation of an instance of the data model.
LIST+EUT	*OK*	DCM-GEN-ABS-01	Levels of abstraction	SHOULD	The data model must abstract from the specific systems, applications and components under attack. Each of them will be part of different contexts (honeypots, real infrastructures, simulations, etc.), and this information should not influence the data model of the attack or threat. The data model must clearly represent the different attack models, tactics or strategies, without being affected by the location or characteristics of the systems from which the data have been gathered.	The semantics used in the data model will not include any meta-data corresponding to information from systems related to real organisations or infrastructures.

Author	Status	ID	Name	Priority	Description	Evaluation
LIST+EUT	*OK*	DCM-GEN-UND-01	Easily understandable by human and security technologies	MUST	The data model must be measurable, precise and easily interpreted by humans and computers. The attack models represented by the data must be validated by cybersecurity specialists. Subsequently, these attack models will be managed and processed by different learning algorithms in order to extract knowledge and intelligence regarding the organised groups that are the source of the different attacks.	The data model must be structured in order to be easily interpreted, with the necessary meta-data to include the necessary information that represents an attack or threat.
LIST+EUT	*OK*	DCM-GEN-REL-01	Designed for broad use/relevance for various domains	SHOULD	This restriction is closely related to the Extensibility requirement (GEN-EXT-01), since the data model must be able to represent any type of attack or threat in any area or domain. Initially, it will have a domain scope, but, as its use is applied in other domains, the data model can evolve to be applied in them.	The data model allows the use of values from different domains without affecting the ability to understand the semantics of the standard part of the data model. The data model allows you to define new data structures (required for different domains), and to link these data structures to the existing data model without affecting the ability to understand the semantics of the standard part of the data model.
LIST+EUT	*OK*	DCM-GEN-OS-01	Open-source	SHOULD	It is recommended that the data model will be Open Source, so its use will be massively extended, and the volume of information that can be gathered will increase considerably.	

Author	Status	ID	Name	Priority	Description	Evaluation
LIST+EUT	*OK*	DCM-GEN-SER-01	Serialisation	COULD	Initially, the data model does not need to be serializable. There is no need for this, since data processing does not require the application of any learning algorithm based on stream processing. However, considering the future, it is not discarded that at any moment a stream processing may be required.	
LIST+EUT	*OK*	DCM-GEN-SHA-01	Sharable	SHOULD	The data model should be easily shared with different entities that may require attack and threat models. These organisations will have in their systems the necessary algorithms to be able to process the shared information. Similarly, the knowledge/intelligence obtained in the processing of the data may be shared.	The data model will use one or more of the attack and threat information sharing standards.
LIST+EUT	*OK*	DCM-GEN-EXP-01	Self-explanatory	COULD	The data model should be self-explanatory concerning threats and attacks they represent. It must contain sufficient information, easily understandable to be processed appropriately, and be able to generate the necessary knowledge/intelligence about organised malicious groups.	The data model must be structured in order to be easily interpreted, with the necessary meta-data to include the necessary information that represents an attack or threat.
LIST+EUT	*OK*	DCM-GEN-CONT-01	Self-contained	SHOULD	This requirement is strongly linked to the self-explanatory one (GEN-EXP-01), where the data model that represents an attack or threat contains all the necessary information for the processing stage, a stage where knowledge/intelligence is generated in relation to malicious organised groups.	The data model must be structured in order to be easily interpreted, with the necessary meta-data to include the necessary information that represents an attack or threat.

Author	Status	ID	Name	Priority	Description	Evaluation
LIST+EUT	*OK*	DCM-GEN-ANO-01	Anonymisation	MUST	The data model should not contain any information that could be used to identify the origin of the information, who has suffered the attack. The confidentiality of the information must be safeguarded. Compliance with this requirement will encourage the deployment and sharing of attack and threat information, helping to generate knowledge/intelligence.	The semantics used in the data model will not include any meta-data corresponding to information from systems related to real organisations or infrastructures. Thus, an attack or threat cannot be related to the organisation/company that has been attacked, or is under potential threat.
KTU	*OK*	DCM-GEN-VAS-01	Visualisation	SHOULD	Visual Analytics is a complex section of data analysis that focuses on the use of information visualisations. The model must define what areas are to be represented. In order to display information, the method must form visualisation objects and transfer them to another task. The model must define the attributes of each object of use and their quantity. The model must include an analytical part. This is which information fields can be changed depending on the information overlap. It must be provided which attributes will be fixed and not changed, and which can be changed by the analyst. What standard will be used for the exchange of information must be described.	

Author	Status	ID	Name	Priority	Description	Evaluation
CNR	*OK*	DCM-INFSHAINT-SEC-01	Data Access Control	MUST	Data protection is a fundamental aspect of a data-sharing platform, also being the main factor for encouraging users to share their data through such a platform. As a matter of fact, users would be encouraged to share their data if they can pair their privacy and security preferences to such data. To this aim, **the data model must allow the data producers to incorporate their security and privacy preferences directly in the data**. As a matter of fact, each piece of data could have security and privacy constraints that are different from all the other pieces of data. For this reason the data model must define a proper fields and a proper format to host such information, because these constraints must be embedded directly in the data. Data protection must be implemented in our data sharing infrastructure by regulating the access and the usage of such data through the adoption of proper access control systems. Access control policies are meant to evaluate whether a request of a subject to perform an action on a given object (objects are data in our case) can be performed. As an access control model, we take into account the Attribute Based Access Control (ABAC) one, because it is very flexible and expressive. This model allows expressing the policies in terms of (even complex and customized) conditions on a set of attributes describing the features of subjects, data and environment. Hence, the data model must include a proper field to embed the attribute	The data-sharing infrastructure allows to write a policy which includes access control conditions, and to embed these policies in the data. These policies are enforced by the data-sharing infrastructure security support when an access/analytic is requested, and the right to access or to execute the analytic could be denied.

Author	Status	ID	Name	Priority	Description	Evaluation
					based access control policy in the data representation.	
CNR	*OK*	DCM-INFSHAINT-SEC-02	Data Usage Control	MUST	Some of the attributes that are taken into account in the conditions that have been listed before for access control (Requirement ID InfShaInt-sec-2) are mutable, i.e., they could change their value over time. Consequently, a condition on such attributes could be satisfied at time T, but could be violated after a while, i.e., at time T+d (where d is a time interval). Hence, the access decision that is taken at time T could be not valid any more at time T+d and, consequently, the right to perform an access that was granted at time T should not be granted any more at time T+d, although the access is still in progress. Usage Control policies can be exploited to express conditions on mutable attributes, that must be continuously verified while the data are in use. This means that **the data model must allow embedding usage control policies in the data,** i.e., it must define a proper field to incorporate usage control policies in the data as well.	The data-sharing infrastructure allows writing a policy which includes ongoing conditions. These policies are continuously enforced by the data-sharing infrastructure security support while the access/analytic is in progress and the accesses/analytics could be interrupted while in progress because of a policy violation due to an attribute change.

Author	Status	ID	Name	Priority	Description	Evaluation
CNR	*OK*	DCM-INFSHAINT-SEC-03	Data Privacy	MUST	Before releasing the data to a subject who requested to read it, or to use the data to perform an analytic, the data producer could want to perform some anonymisation operations on such data, for privacy purposes. As an example, if the data are the logs of some internal services, the data producer could want to anonymise the last digits of the IP addresses in these logs because he does not want to release information about his internal service structure. As specified in requirements InfShaInt-sec-1 and InfShaInt-sec-2, **the data model must allow data producers to embed data manipulation operations in the data**.	The data-sharing infrastructure allows to write a policy which includes anonymisation operations and to embed these policies in the data. These policies are enforced by the data-sharing infrastructure security support when an access/analytic is requested, and the data are anonymised before being released.
CNR	*OK*	DCM-INFSHAINT-SEC-04	Obligations	MUST	Obligations are actions that are executed by the system as a consequence of the decision process. These actions could be performed on the data or could be not related to the data. For instance, the data producer could want to receive an email every time that one of his data objects is read or used to perform an analytic. Another example is the one where a data producer wants that subjects accept a disclaimer before using his data. As specified in requirements InfShaInt-sec-1, InfShaInt-sec-2, and InfShaInt-sec-2, **the data model must allow data producers to embed obligations in the data**.	The data-sharing infrastructure allows us to write a policy which includes obligations and to embed these in the data. These policies are enforced by the data-sharing infrastructure security support when an access/analytic is requested, and the obligations are properly executed.

Author	Status	ID	Name	Priority	Description	Evaluation
CNR	*OK*	DCM-INFSHAINT-ANALY-01	Metadata for Analytics	MUST	The infrastructure for data sharing supports their integration through the execution of collaborative data analytic, i.e., analytic functions that are executed exploiting data produced by several subjects and stored on the data-sharing infrastructure. This considerably increases the value of data, because exploiting data from several sources to return a global result would enable to obtain earlier or better results. Hence, **the data model must allow data producers to specify all the relevant details concerning the format of the data** in order to allow the data-sharing platform to understand for which analytic they can be used as input.	The data-sharing infrastructure integrates an engine for the execution of the collaborative analytic function and it exposes an API or a GUI from which the supported analytic functions can be invoked on the proper set of data.
LEO	*OK*	DCM-SPEC-TIM-01	Capability to crawl data from OSINT	MUST	The system must be able to capture information from social networks (e.g. Twitter, blogs) using RSS and on the anonymous sharing source Pastebin.com.	Simulation of posting a Post in Twitter, the connection of an RSS or Pastebin page to the system and verification that the information entered is captured by the system.
LEO	*OK*	DCM-SPEC-TIM-02	Capability to create specific ontology to capture Cyber Attack Events	COULD	The system must allow the configuration of rules that are created using Boolean operators and pattern matching. It must be possible to configure one or more rules, which can be grouped into scenarios (which are collectors of rules). The collection of rules could be part of the data model.	Configuration of a rule named "Nuclear Power Plant" & "Rome", posting of such keywords on Pastebin and verification that the platform captures the word Nuclear Power Plant associated with the word Rome, following the Boolean

Author	Status	ID	Name	Priority	Description	Evaluation
						rules.
LEO	*OK*	DCM-SPEC-TIM-03	Capability to extract entity	MUST	The system must be able to extract entities from the text captured from open sources such as Names of people, Name of Organisations, locations, IP, CVE, names of hacker operations, hashtags, names of users mentioned, username of whom shared posts and links. The model must support these identifiers as a minimum.	Inserire su una delle fonti aperte sotto controllo, un testo che contenga al suo interno Nomi di Persone: Es. Giovanni Micolucci, Nome di Organiszazioni: lavora presso la Leonardo Company, Locazioni: nella sede di Chieti in Italia, IP: Es. il suo e 172.20.20.20, CVE: la sua macchina e affetta dal CVE-2018-3333, nomi di hacker oepration: che e stata usata nella #ophackleo e Hashtag: Es. #italytrend, nomi di utenti citati nel post: Es. @giovannimicolucci, autore del post: Claudio Porretti e link: Es. nel testo e incluso http://sitotest.com. Insert in one of the open sources a text containing: Person name, Organisation name and location, IP, CVE his machine is affected by, names of hacker operation (which was used in the #ophackleo) and Hashtag, user names mentioned in the post, author of the post, and link

Author	Status	ID	Name	Priority	Description	Evaluation
CESNET	*OK*	DCM-SPEC-DDOS-01	DDoS sample	COULD	The data model provides a mean to include a sample of the network traffic as a sighting.	A pcap file containing network traffic is attached to the data model and the same information contained is also represented by the data model.
CESNET	*OK*	DCM-SPEC-DDOS-02	Spoofed identifiers	COULD	The data model provides a mean to indicate that an identifier might be spoofed (e.g. when an attacker spoofs its source IP addresses).	A user is able to tell if an identifier of an attacker has been spoofed based on the description of a threat/attack using the common data model.
CESNET	*OK*	DCM-SPEC-DDOS-03	Attribution to a botnet	COULD	The data model provides a mean to link particular attack with a particular botnet. Therefore the model is supposed to have an attack categorisation scheme together with the ability to reference external information sources, e.g. with URL.	
CESNET	*OK*	DCM-SPEC-DDOS-04	Mitigation rule	COULD	The data model provides a mean to capture mitigation actions, for example, a rule that can filter out the DDoS traffic.	
CESNET	*OK*	DCM-SPEC-DDOS-05	Attack/threat categories and subcategories	COULD	The data model provides a mean to indicate not only the basic category of an attack (e.g. DDoS) but also its variants (e.g. ddos.synflood, ddos.dnsamplification).	The taxonomy of attacks is fine-grained to the level of attack/threats variants and a user might select either just the category or the category and its subcategories.

Author	Status	ID	Name	Priority	Description	Evaluation
KTU	*OK*	DCM-SPEC-VAS-01	Visual object storing and communicating	SHOULD	The model must describe attributes in the following fields: 1. Storage systems, 2. File systems protected using cryptographic schemes, 3. Cloud-backed file storage, 4. Communication The model must describe information objects from the following areas: 1. IT systems subject to visual analysis; 2. vulnerabilities present on those systems; 3. how cyber-attacks are exploiting such vulnerabilities; 4. the actual impact of those attacks on the services and goals offered on the IT systems analysed.	What information will be aggregated and from which areas
KTU	*OK*	DCM-SPEC-VAS-02	Visualisation open-source	SHOULD	The model must have an ontology or taxonomy that allows the selection of adequate tools for the sharing of visualisation information. The model must offer a list of open-source tools.	Ontology or Taxonomy
NASK	*OK*	DCM-SPEC-MALW-01	Malware families	SHOULD	The data model provides a way to refer to malware families. The primary taxonomy of families is based on Malpedia (https://malpedia.caad.fkie.fraunhofer.de/families), however additional families can be specified.	Describe a malware sample by linking it to a known malware family.

Author	Status	ID	Name	Priority	Description	Evaluation
NASK	*OK*	DCM-SPEC-MALW-02	File properties	SHOULD	The data model allows describing a malware sample using a set of attributes from static analysis. In particular, the following set of attributes is supported: file size, file type, cryptographic hashes (MD, SHA-1, SHA-256, SHA-512), fuzzy hashes (ssdeep, TLSH). Other attributes can be added when available.	Describe a malware sample using a set of attributes.
NASK	*OK*	DCM-SPEC-MALW-03	Malware configuration	COULD	The data model includes a way to express static configuration extracted from malware samples and dynamic configuration obtained from command and control servers. The configuration can be structured (key-value dictionary supporting nesting) or in the form of a binary blob (typically unstructured text). There is a way to link static configurations to malware samples that they were extracted from and dynamic configurations to static configurations that are used to establish a connection to the command and control server.	Describe an entire chain of malware analysis: original sample, extracted static configuration and dynamic configurations from the command and control server, preserving the parent-child relationships between entities.

Author	Status	ID	Name	Priority	Description	Evaluation
NASK	*OK*	DCM-SPEC-MALW-04	Decompile	MUST	The data model includes entities corresponding to functions extracted from malware binaries. The functions can be in the form of decompiled source, disassembled or a generalized version of thereof (canonical version). The data model also provides a way to express the relationship between functions and samples that they are part of and between function (similar function, identical function, canonical vs concrete implementation).	Describe the relationship between malware samples based on the co-occurring functions.
NASK	*OK*	DCM-SPEC-MALW-05	Malware similarity	MUST	The data model allows expressing similarity between a pair of malware samples based on a different method. Degree of similarity is expressed as a numeric value (0-100) and a method used to establish the degree of similarity is provided.	Describe the similarity between two samples.
NASK	*OK*	DCM-SPEC-MALW-06	Malware clusters	MUST	The data model allows marking a malware sample as belonging to one or more clusters. A cluster is identified by a unique alphanumeric label.	Add two samples to the same cluster.
LMT	*OK*	DCM-INFSHAINT-SEC-05	Data Access Control	MUST	The data model must have a safeguard in place not only to protect which partner can access which data (InfShaInt-sec-1), but also limit data model owner/administrator from accessing data added to the data model. Some	The data-sharing infrastructure limits platform administrators to view or extract published data.

Author	Status	ID	Name	Priority	Description	Evaluation
					form of expressing data classification (e.g. distribution level, TLP) from an access perspective.	
LMT	*OK*	DCM-GEN-EXT-02	Extendibility	MUST	The data model must support data processing in Nodes, located at Partner locations. Nodes should support data manipulations (InfShaInt-sec-3) at Partners premises, before leaving for future analyses.	Data sharing infrastructure supports node system, that can be installed at partners' locations and used to extract/process data
LMT	*OK*	DCM-GEN-DEL-01	Data withdraw	MUST	There should be an option to withdraw data published to the data model or render them invalid.	Data sharing model allows withdrawing existing data from data-sharing platform.
L3CE	*OK*	DCM-DATA-INT-01	Cross-integrity between cyber-physical incidents, cyber-information incidents, physical-information incidents and strategic events	MUST	Data model must ensure cross-integrity between cyber-physical incidents, cyber-information incidents, physical-information incidents and strategic events in the following meta attributes in order to ensure further analytic ability: - Event/Incident Timestamping: timestamp of origination; - Event/Incident Timestamping: timestamping of first detection; - Event/Incident Timestamping: timestamping the effect/impact start; - Event/Incident Timestamping: timestamping the effect/impact end; - Source identification; - Geo location of origin; - Geo location of incident; - Geo location of effect/impact; - Author or source or adversary IP; - Category (cyber, information, event); - Target;	

Author	Status	ID	Name	Priority	Description	Evaluation
					- Relation to known adversaries (criminal, actors, groups, adversaries);	
L3CE	*OK*	DCM-DATA-INT-02	Data extensions, attachments	MUST	The data model must support extensions, attachments of original source info objects, would it be a cyber incident or social media activity. As all sources maintain their data structures, the extensions/attachments should be embedded via original source structure (Facebook, Twitter, VK, publications)	
L3CE	*OK*	DCM-DATA-INT-03	Link support	MUST	The data model must support link to ongoing strategic process: elections stage in T-Shark umbrella case or other strategic event stages.	

Chapter 11 Annex – B: Raw combined SWOT Matrix

Table 11.1: Combined SWOT

Internal		
ID	**Strengths**	**Weakness**
NIST Cybersecurity Framework	Straight guidelines and recommendations.	Relies on third party standards, certifications and guidelines
	Nice starting point for further development.	Not meant to be a single solution, but a complement to other corporative processes and protocols
	Applies to organisations of all sizes	CCE is still not part of the framework. It is advisable, to integrate the standard into it.
	Covers the whole spectrum of organisations, from small SMEs to Critical Infrastructures.	Assessment depth open to organisational preferences. No control requirements.
	Flexible and adaptable.	The framework doesn't measure risk
	Open, free initiative. Nicely maintained.	No level guidance for companies.
	Versatile.	There is no focus on any of the financial aspects.
	Apply to organisations of all sizes.	Due to the plethora of documentation, it becomes quite complex for the organisations/ individuals to find what they are looking for.
	Customisable according to the nature of the organisation.	The NIST Framework focuses only on how to plan and implement IT security, but not on the entire information management system.
	It is designed to be applicable regardless of the requirements and the technologies.	
	Widely recognised standard.	
	High-level guidelines to translate them according to specific needs and strategies.	
	Applicability to public and private organisations.	
	Very well structured.	
NATO AC/35	General and flexible, may be compatible with other methodologies and tools.	Not specific enough, it does not prescribe a specific management method or tool.

	Extensive guidance on how risk assessment fits within the system development lifecycle and how it needs to be executed.	Focus on military systems as it was prepared by NATO.
	It is suitable to every methodology or tool since it does not prescribe a specific risk management method or tool.	It is built mainly for military purposes so its use may be limited.
	Military "field-tested' methodology.	Separate risk management guidelines have been published in many NATO nations.
	Very structured risk assessment and risk management guidelines for military systems.	Not a widely accepted/known standard.
		Does not prescribe a general security management method but it focuses on risk management.
		Not generic but specific for military systems
EBIOS	Compatible with international standards.	Requires a significant level of expertise in security analysis
	Totally open, with strong support and tools.	Lack of audit and continuous integration methods
	Exhaustive approach for risk analysis.	The last phases are too theoretical and difficult to implement
	It is often revised.	In EBIOS, it is necessary to define security requirements which may be difficult to be determined in the early stages of a project.
	It totally complies with the latest ISO standards.	The method is somewhat complex since modules decomposed in activities and activities in actions.
	Compatibility with international standards	The EBIOS method does not provide immediate solutions to security problems, but gives only support
	Comprehensive approach: the structured procedure of the EBIOS method allows identifying and combining the constituent elements of the risks.	
	Context adaptability	
MITRE Att&ck	Powerful and complete taxonomy on cyber threats. Deep and rich in technical details.	Not a standalone solution; it is designed to complement other tools, operations and processes, but it is not one by itself.
	Provides deep understanding of ATPs and TTPs.	Some degree of ambiguity demands further interpretation.
	Largely maintained and updated to	Classifications may be incomplete or

	the SoA of cyber threats.	complex to fit-in exactly on one term of the taxonomy.
	A wide set of technical tools, guidelines and documentation.	Guidelines are flexible and informal.
	Flexible.	Too technical.
	Describes attack from the attacker point of view, provides knowledge of the attacker and its profile.	Very complex, with a lot of information and a lot of attack patterns
	Provides advice, guidance and potential countermeasures.	• The participation of security specialists is necessary for the definition of models of attack tactics and strategies. • It is very complete and very complex, since the possibilities of relations between the different entities are very wide.
	The technical description that covers a lot of attack scenarios and combination.	The framework does not emulate any legitimate applications or processes that have nothing to do with the attacks.
	• Flexible • Describes the tactics and strategies employed by attackers. • It includes controls and countermeasures for each of the tactics and strategies described. • Allows to model attacks based on defined tactics and strategies. • Versatile: information systems and ICS.	Very complex and technical.
	Provides significant knowledge, as is approaches the attacker's perspective.	Specifically related to adversary attacks.
	Provides a fully detailed approach based on a user-friendly matrix presentation.	
	It is flexible and covers a wide range of levels (high to low level).	
	The knowledge base of adversary tactics and techniques based on real-world observations of cyberattacks.	
	The modern way of looking at cyberattacks. Based on tactics and techniques that indicate an attack is in progress.	
ISO 2700X	Industry Standard. Widely used. It allows security professional to check whether all information security gaps are covered, in an ISM. That is why, any information security framework should be compliant with ISO2700	It is a too high level, and enterprise oriented standard, it addresses security from a managerial point of view.

	family.	
		The level of security must be defined by the user.
	Apply to all types of organisation.	Not very technical, high level framework.
	Flexible, allows the user to select a method, or more likely several methods and/or tools, that suit their organisation's requirements.	Focused on Business to Business.
	Broad in scope.	Time-consuming (vague and not easy to understand).
	Continually evolves, regularly updated to remain relevant.	The context definition, scope and risk acceptance are defined by the user.
	It is the most used and widely recognised standard.	Does not provides a specific methodology to implement the requirements.
	It can be implemented in any kind of organisation.	The standard is too generic, it does not go into enough detail.
	Very Flexible and solid.	It mostly serves Business-to-Business purposes.
	World-renowned security guideline.	Profiles cannot be used to establish minimum requirements for other organisations, such as suppliers or partners.
	Simple methodological structure	
	Ability to identify, analyze and deal with an organisation's information risks to protect itself from cyber threats and data breaches.	
	Applicability to all types of organisations, public or private, profit or non-profit, regardless of size or industry.	
	Certification.	
DDoS_mirkovic	Has a capability to reference external information (e.g. in the form of a URL to the description of a botnet).	High specificity for particular DDoS subcase.
	Has a capability to capture mitigation actions.	High complexity to capture all the phases of an incident including the mitigation and recovery phase.
STIX & TAXII	• Flexible. • Data model is built upon eight principal concepts. • Can be used for multiple cases of use. • The standard has been widely recognised by the community and the number of tools compatible with STIX is growing.	• STIX and TAXII might be difficult to use and comprehend at first.

CybOX	• Flexible. • The goal of CybOX is to provide a common structure to represent cyber observables . • It complements and links to different MITRE standards. • It is also widely recognised by the community as a complete and well-developed taxonomy for the representation of cyber incidents.	• High granularity that makes implementation difficult.
VERIS	• Flexible • Data model for defining and exchanging incident information. • Introduces categories and metrics to describe incidents.	• Verizon has developed his own solution for storing and describing incidents. Although the VERIS taxonomy can be used without the Verizon solution, it is often coupled with it in practical use.
OpenIOC	• Flexible • Defined language for handling forensic information. • Extensible format by defining new data types.	• OpenIOC is useful for the representation of indicators of compromise, mostly used for detection in software. It is, however, less suitable for information exchange since it was not created for this purpose.
AVOIDIT	• AVOIDIT aims to help identify and defend against cyber-attacks. • AVOIDIT taxonomy represents an interesting way of linking multiple elements of an attack.	• This taxonomy has not yet been implemented in any information sharing tool and has not had a large response from the community.
CAPEC	• Flexible. • CAPEC entries are descriptions of particular attack patterns, that is, the techniques and procedures used to carry out the sequence of steps that makes up the pattern. • The standard has been widely recognised by the community and the number of tools compatible with STIX is growing.	• The participation of malware and attack analysis specialists is necessary to build the model of a given attack.
MISP	Different levels and types to express threat, event or incident: Events, Objects, Object References, Tags, Sightings, MISP Galaxy.	MISP was originally designed for malware.
	Can be used to share technical and non-technical information about malware samples, incidents, attacks and general intelligence	MISP official taxonomies do not currently meet some requirements.
	Open-source.	
	JSON format.	
STIX 2	- The development follows a formalized and open process - JSON-based – easy to parse (syntactic level) - Wide scope: describes most types	- Needs extensions to accommodate more specific use cases - Graph model requires more effort to process on the semantic level (entities and their relationships)

	of entities that are shared in practice - Flexible: relationship graph - Extendible: new entities, values, attributes and relationships are straightforward to add - Includes existing taxonomies	
MAEC	- Rich vocabularies - Extendible (entities, vocabularies and relationships)	- Focus on behavioural analysis, limitations concerning results of static analyses - Not general-purpose (only malware)
IntelMQ (Data Harmonisation Ontology)	- Simple to parse - JSON-based (syntactic level) - Simple to interpret and process – flat list of well-defined attributes (semantic level)	- Primarily designed for remediation feeds (abuse notifications), not-general purpose - Difficult to add data with more structure (beyond key-value)
n6	- Simple to parse JSON-based (syntactic level) - Simple to interpret and process – flat list of well-defined attributes (semantic level)	- Primarily designed for remediation feeds (abuse notifications), not-general purpose - Limited documentation
CVE	- Simple model, easy to interpret and parse	- Limited in scope: just identification of vulnerabilities
TLP	- Very simple, easy to use	- Limited in scope: just classification of information - No relation to formal classification schemes
CVSS	- Simple model, easy to parse - Relatively easy to interpret	- Limited in scope: just rating of severity of vulnerabilities - Interpretation of scores is subjective
ENISA Reference Security Incident Taxonomy	- Simple structure - Covers all types of common incidents - Rich subtypes	- Scope limited only to incident taxonomy