Advanced Cyber-Threat Intelligence, Detection, and Mitigation Platform for a Trusted Internet of Things

Grant Agreement: 786698

D2.2 Threat sharing methods: comparative analysis

Work Package 2: Cyber-threat landscape and end-user requirements

Document Dissemination Level

PU	Public	X
CO	Confidential, only for members of the Consortium (including the Commission Services)	

Document Due Date: 31/08/2018

Document Submission Date: 29/08/2018

Co-funded by the Horizon 2020 Framework Programme of the European Union

Document Information

Deliverable number:	D2.2
Deliverable title:	Threat sharing methods: comparative analysis
Deliverable version:	1.00
Work Package number:	WP2
Work Package title:	Cyber-threat landscape and end-user requirements
Due Date of delivery:	31/08/2018
Actual date of delivery:	31/08/2018
Dissemination level:	PU
Editor(s):	Spiros Skiadopoulos (UOP)
Contributor(s):	Vasiliki-Georgia Bilali (KEMEA) Sotirios Brotsis (UOP) Thanasis Chatzios (UOP) Dimitris Kavallieros (KEMEA) George Kokkinis (KEMEA) Nicholas Kolokotronis (UOP) Paris Koloveas (UOP) Andrew Ramsdale (CSCAN) Stavros Shiaeles (CSCAN) Spiros Skiadopoulos (UOP) Christos Tryfonopoulos (UOP)
Reviewer(s):	Emanuele Bellini (MATHEMA) Liza Charalambous (ADITESS)
Project name:	Advanced Cyber-Threat Intelligence, Detection, and Mitigation Platform for a Trusted Internet of Things
Project Acronym	Cyber-Trust
Project starting date:	01/05/2018
Project duration:	36 months
Rights:	Cyber-Trust Consortium

Version History

Version	**Date**	**Beneficiary**	**Description**
0.10	15/05/2018	UOP	Proposed outline
0.20	07/06/2018	UOP	First draft of Section 3
0.30	18/06/2018	UOP	First draft of Section 7
0.40	13/07/2018	UOP, KEMEA, CSCAN	First draft of Section 5
0.50	30/07/2018	KEMEA	First draft of Section 6
0.60	31/07/2018	UOP	First draft of Section 2
0.70	03/08/2018	CSCAN	First draft of Section 4
0.80	06/08/2018	UOP	Sections 1-3 and 7-12 are final

0.90	07/08/2018	UOP	Sections 4-6 are final
1.00	08/08/2018	UOP	Deliverable is ready to be reviewed
1.10	19/08/2018	UOP	Review comments implemented in deliverable
1.20	24/08/2018	UOP	Final document produced

Acronyms

ACRONYM	EXPLANATION
API	Application Programming Interface
AV	Anti-Virus
CAPEC	Common Attack Pattern Enumeration and Classification
CAPTCHA	Completely Automated Public Turing test to tell Computers and Humans Apart
CCE	Common Configuration Enumeration
CCSS	Common Configuration Scoring System
CERT	Computer Emergency Response Team
CISO	Chief Information Security Officer
COA	Course of Action
CPE	Common Platform Enumeration
CSIRT	Computer Security Incident Response Teams
CTI	Cyber-Threat Intelligence
CVE	Common Vulnerabilities and Exposures
CVRF	Common Vulnerability Reporting Framework
CVSS	Common Vulnerability Scoring System
CWE	Common Weakness Enumeration
CWSS	Common Weakness Scoring System
CYBEX	Cybersecurity Information Exchange Framework
CybOX	Cyber Observable Expression
DDoS	Distributed Denial of Service
DHS	Department of Homeland Security
DNS	Domain Name Server
DPI	Deep Packet Inspection
HIDS	Host-based Intrusion Detection System
IDS	Intrusion Detection System
IoC	Indicator of Compromise

ACRONYM	EXPLANATION
IODEF	Incident Object Description Exchange Format
IoT	Internet of Things
IPS	Intrusion Prevention System
IR	Incident Response
ISAC	Information Sharing and Analysis Center
ISAO	Information Sharing and Analysis Organization
IUO	Investigational Use Only
JSON	JavaScript Object Notation
LEA	Law Enforcement Agency
MAEC	Malware Attribute Enumeration and Characterization
MSM	Making Security Measurable
NAC	Network Access Control
NIDS	Network Intrusion Detection System
NLP	Natural Language Processing
NSM	Network Security Monitoring
OCIL	Open Checklist Interactive Language
OS	Operating System
OSINT	Open Source INTelligence
OVA	Open Virtual Appliance
OVAL	Open Vulnerability and Assessment Language
PII	Personally Identifiable Information
PoS	Point of Sale
pub/sub	publish/subscribe
REST	Representational State Transfer
RID	Real-time Inter-network Defense
SaaS	Software as a Service
SDK	Software Development Kit

ACRONYM	EXPLANATION
SIEM	Security Information and Event Management
SOC	Security Operations Center
STIX	Structured Threat Information Expression
TAXII	Trusted Automated eXchange of Indicator Information
TIP	Threat Intelligence Platform
TOR	The Onion Router (the dark web browser)
TTPs	Tactics, Techniques and Procedures
UI	User Interface
URL	Uniform Resource Locator
XCCDF	Extensible Configuration Checklist Description Format

Table of Contents

List of Figures

List of Tables

Executive summary

Organizations worldwide, from governments to public and corporate enterprises, are under constant threat by evolving cyber-attacks. The fact that there are literally billions of IoT devices globally, most of which are readily accessible and easily hacked, allows threat actors to use them as the cyber-weapon delivery system of choice in many today's cyber-attacks, e.g., from botnet-building for launching distributed denial of service attacks, to malware spreading and spamming. The sooner an organization knows about emerging threats, the more efficiently cyber-defense mechanisms will be utilized. Therefore, the main challenge organizations face is the abundance of data and the lack of actionable intelligence.

Cyber-threat intelligence is any information that can help an organization identify, assess, monitor, and respond to cyber threats. Examples of such information include indicators (system artifacts or observables associated with an attack), security alerts, threat intelligence reports, as well as recommended security tool configurations. As most organizations already produce an enormous amount of cyber-threat information in multiple forms and types, it is crucial for effective cyber-defense to share both internally and externally the available data as part of their information technology and security operations efforts. The goal of the work carried out, and reflected in this deliverable, is to identify best practices in this area. Disseminating the details of identified vulnerabilities amongst the cyber-security experts, verifying their legitimacy (i.e., that they indeed pose a threat), and rating their impact is critical. This deliverable overviews and critically evaluates existing industry-wide vulnerability reporting and sharing sources, standards, frameworks and platforms in order to provide recommendations on the approach to be followed in the Cyber-Trust platform.

We begin by presenting the methodology of our analysis. Then, we review several data sources for threat information sharing systems categorized into internal, community, and external with the purpose of compiling a cataloging inventory that contains elements useful for the purposes of the project. Such elements include the type of exposed data (e.g., structured machine-readable or unstructured) and query languages, protocols, or services available for data retrieval.

Subsequently, we consider and report the appropriateness of different vulnerability frameworks for disseminating the identified cyber-threats across different organizations and promoting awareness about emerging cyber-threats. Moreover, issues pertaining to the basic structure, the key elements (i.e., expressiveness, flexibility, extensibility, automation, structuring), and prominent strengths/weaknesses of the presented frameworks are discussed and critically evaluated within the scope of the Cyber-Trust project. Frameworks and languages for supporting expressive content-based subscriptions in the context of specialized pub/sub services for cyber-threat information push are also considered.

Following, we illustrate how the presented frameworks and languages are realized in platform and tool implementations to provide the necessary functionality and enhance standard adoption. The mechanisms for handling structured cyber-threat information for a wide variety of use cases (including those outlined in the project) are also presented alongside important components that include the key characteristics of each platform, the supported observables and schemas, and the adopted standards.

Next, we review several prominent market solutions related to the discovery and management of cyber threat intelligence and categorize them into services, data feeds, platforms, and complete systems. The main features and characteristics with respect to a number of different facets -including architecture, offered services, standards' adoption, and mode of operation- are critically compared for each category to highlight salient market practices that relate to the goals of the Cyber-Trust project.

Finally, based on our analysis, we present our recommendations for the Cyber-Trust project. In a nutshell, we propose to use STIX as the sharing mechanism and MISP as the sharing platform.

The present deliverable provides a comparative analysis of threat sharing methods. Thus, it is quite technical by nature. We believe readers with technical knowledge (such as the personnel of LEA, ISAO, ISAC and IPA) will find our presentation comprehensive and our analysis accurate and complete. Non-technical readers might have to skip the engineering parts (especially during their first reading).

1. Introduction

Cyber-threat intelligence (CTI) is any information that can help an organization identify, assess, monitor, and respond to cyber threats. Examples of such information include indicators (system artifacts or observables associated with an attack), security alerts, threat intelligence reports, as well as recommended security tool configurations. As most organizations already produce an enormous amount of cyber-threat information in multiple forms and types, it is crucial for effective cyber-defense to share both internally and externally the available data as part of their information technology and security operations efforts. The goal of the work carried out, and reflected in this deliverable, is to identify best practices in this area.

1.1. Purpose of the document

Disseminating the details of identified vulnerabilities amongst the cyber-security experts, verifying their legitimacy (i.e., that they indeed pose a threat), and rating their impact is critical. The purpose of the document is to overview and critically evaluate existing industry-wide vulnerability reporting and sharing sources, standards, frameworks and platforms in order to provide recommendations on the approach to be followed in the Cyber-Trust platform.

Initially, sources of data for threat information sharing systems are presented and categorized into internal, community, and external with the purpose of compiling a cataloging inventory that contains elements useful for the purposes of the project. Such elements may include the type of exposed data (e.g., structured machine-readable or unstructured) and query languages, protocols, or services available for data retrieval.

Subsequently, the appropriateness of different vulnerability reporting frameworks for disseminating the identified cyber-threats across different organizations and promoting awareness about emerging cyber-threats is considered. Moreover, issues pertaining to the basic structure, the key elements (i.e., expressiveness, flexibility, extensibility, automation, structuring), and prominent strengths and weaknesses of the presented frameworks are discussed and critically evaluated within the scope of the Cyber-Trust project. Frameworks and languages for supporting expressive content-based subscriptions in the context of specialized pub/sub services for cyber-threat information push are also considered.

The presented frameworks and languages are realized in platform and tool implementations to provide the necessary functionality and enhance standard adoption. The mechanisms for handling structured cyber-threat information for a wide variety of use cases (including those outlined in the project) are also presented alongside important components that include the key characteristics of each platform, the supported observables and schemas, and the adopted standards.

Finally, prominent market solutions related to the discovery and management of cyber threat intelligence are reviewed and categorized into services, data feeds, platforms, and complete systems. The main features and characteristics with respect to a number of different facets including architecture, offered services, standards’ adoption, and mode of operation are critically compared for each category to highlight salient market practices that relate to the goals of the Cyber-Trust project.

1.2. Relations to other activities in the project

The reviewed data sources, standards, frameworks and platforms in the present document are directly related to a number of different tasks within the Cyber-Trust project; data source identification, standard adoption and framework support for the chosen standards are directly related to Tasks T2.2 - Use case scenarios and T2.3 - End-user requirements and specifications, while platform and market analysis for the most prominent solutions are pertinent to Task T5.1 - Threat intelligence techniques.

Specifically, in Task T2.2 a number of representative use cases will be defined in order to demonstrate the full potential of the Cyber-Trust approach in detecting and mitigating advanced cyber-attacks. To this end, the present document reviews state-of-the-art data sources, standards, and frameworks and examines their suitability within the context of the identified use cases, by resorting to several criteria (e.g., expressiveness, flexibility, extensibility, automation) of each solution. Moreover, in Task T2.3 the needs and requirements of the end-users regarding the detection and reaction to advanced cyber-attacks will be identified and analyzed. To this end, the overview and critical evaluation of the different frameworks will assist the process of understanding end-user needs in terms of best practices and technologies, while the recommendations provided will maximize the project's operational efficiency with respect to the users' specific needs.

Similarly, cross-comparison of available platforms and market analysis are important inputs for Task T5.1 that aims at gathering public cyber-threat intelligence information and leveraging it to identify emerging threats. The architectural overview of existing solutions will shape the tools and services to be used, whereas the standard and framework recommendations will be taken into account for representing cyber-threat information and sharing it to different stakeholders (i.e., devices, users, organizations).

1.3. Structure of the document

The document is comprised of ten sections, the first being the current introductory section. The rest of the document is organized as follows. Section 0 presents an overview of cyber-threat intelligence by providing definitions, categorizations, and uses of CTI, while also presenting the key CTI sharing benefits and challenges. Subsequently, Section 3. Description of methodology outlines the methodology used for providing recommendations on the CTI sharing aspects of the Cyber-Trust project by relying on a set of high-level requirements. Section 4. CTI sources identifies available CTI sources, organizes them according to their type to internal, community and external, and highlights key elements that are of interest to the project. Section 5. CTI formats and languages focuses on efforts to standardize data formats and exchange protocols related to CTI, while Section 6. CTI platforms and tools presents tools and platforms and critically compares the available solutions against key elements (e.g., expressiveness, flexibility, extensibility, automation, structuring) and characteristics (e.g., supported observables, schemas, adopted standards). Section 7. Current market situation analyzes the current market situation by reviewing several market solutions related to the discovery and management of CTI and categorizing them according to the provided services. Section 8. Recommendations provides recommendations regarding specific standards, languages, platforms and tools to be adopted by the Cyber-Trust project, while Section 9. Conclusions concludes the deliverable. Finally, the bibliography in Section 10. References is followed by Annex A that links shared CTI with issues related to information privacy and sensitivity, and Annex B that discusses the traffic light protocol for specifying restrictions on CTI information.

2. CTI sharing overview

Cyber-threat intelligence is the part of classic intelligence that relates to networks, computers, and other types of information technology. Intelligence is the information and knowledge that is gained about an adversary by means of observation and analysis; alternatively, it could be described as being actionable information for dealing with an adversary. There are two significant points derived by the previous views that could be applied to CTI. Firstly, intelligence is not just data, but information that has been analyzed, and secondly intelligence must be actionable for various reasons. Furthermore, cyber threat intelligence can be categorized as tactical or strategic. Strategic intelligence is composed of things like how to motivate the adversaries and tactical intelligence is composed of things like *tactics, techniques and procedures* (TTP) and *indicators of compromise* (IoCs). IoCs can be considered as one of the most actionable types of CTI and are often the most valuable standards and tools. IP addresses, uniform resource locators (URLs), file hashes and domain names are some of the most commonly used IoCs.

Over the decades cyber threats have grown, morphed and become more sophisticated. Adversaries may now use a vast set of tools and tactics in order to attack their victims with their motivations ranging from intelligence collection to destruction or financial gain. Understanding the attacker nowadays has gotten more important, but not less complicated. The way to identify and understand an attacker as well as the use of that information to protect networks is a fundamental concept behind cyber threat intelligence. Threat intelligence is focused on the analysis of the capabilities, motivations, and the goals of an adversary; and CTI is focused on how these goals are achieved using the cyber domains.

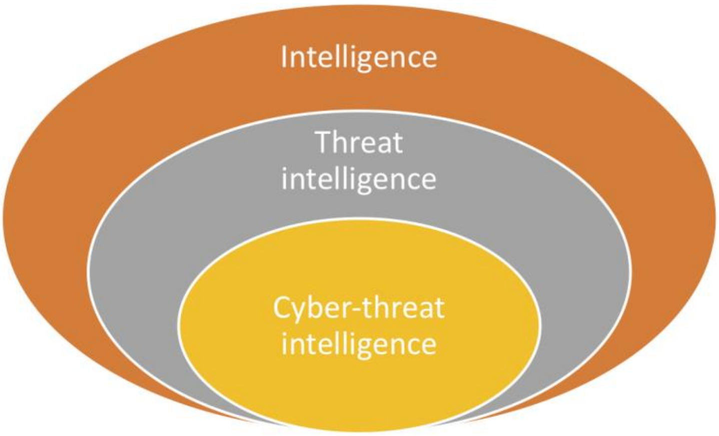


Figure 1. From intelligence to cyber-threat intelligence

Information security analysts usually focus on scientific concepts that are testable and reproducible, but there is art and science behind cyber threat intelligence. Art includes the analysis and the interpretation of data about the attackers and how to disseminate that information to the audience in an easier way in order to provide incentives to act. This will help security analysts to understand a thinking, evolving and reacting adversary.

Security analysts are equipped with technologies to identify malicious activity on networks and are able to trace that maliciousness back to the attacker. Over the years, these new technologies were developed to aim in increased detectability of the malicious activity on networks (DPI, NAC and network security intelligence appliances etc., which are based on familiar concepts with new applications).

2.1. Threat information types

As cyber threat is defined as *any circumstance or event with the ability to adversely impact individuals, organizational operations, organizations, or any nation through an information system by means of unauthorized access, modification or destruction of information or denial of service*. As *threat actors* or simply actors we define individuals and groups that are posing threats.

Cyber threat information is any kind of information that could help an organization to protect itself against any threat and detect the activities of an actor. There are many types of threat information that include the following:

- **Indicators.** These are observables or technical artifacts suggesting that an attack is going to happened or that a compromise of the system has already occurred. In other words, indicators can be used in a system to protect it against any potential threats. Examples of indicators include the IP address of a suspected command, a distrustful DNS domain name, a URL that references suspicious content, a file hash using a malicious executable, or text code of a malicious email message.
- **Tactics, techniques, and procedures.** TTPs are used to describe the behavior of an actor; tactics are descriptions of behavior, techniques are descriptions of tactics, while procedures are detailed descriptions in the context of a technique. TTPs describe the willingness of an actor to use a specific attacking tool, a malware variant, an exploit or a delivery mechanism (e.g., phishing).
- **Security alerts.** Also known as advisories, security alerts are brief and human-readable notifications regarding vulnerabilities, exploits or any other security issue.
- **Threat intelligence reports.** These include documents that describe TTPs, types of systems, actors and target information, and any other information related to cyber threats that provide greater awareness to an organization.
- **Tool configurations.** These include recommendations for the installation and the use of mechanisms in order to collect, process, exchange and analyze threat information. Tool configuration information may consist of instructions on how to customize and use intrusion detection signatures, web filter configuration files, or firewall rules.

The production and sharing of threat information internally is been adopted by many organizations. For example, a security team may identify compromised files on a system and produce an associated set of indicators (e.g., file names, hash values), which can be shared with system administrators who configure security tools (such as HIDSs), in order to prevent or detect their presence on other systems. The goal is to support threat information sharing practices between organizations, and to provide internal, threat information to other organizations.

2.2. Sources and methods

Traditional intelligence (INT) sources are most often centered on the INTs, which describe where the data is collected from (*see* e.g., [65]):

- **HUMINT.** Human-source intelligence as the oldest form of intelligence collection is derived from humans. There is a serious concern about whether cyber threat intelligence could be derived from HUMINT. An example is to conversate with individuals who have knowledge of intrusions. Many people describe as HUMINT all the information gained from communication and interactions with individuals by means of restricted online forums, but this kind of information gathering could also be considered as SIGINT, because it is emanated from electronic communications.

- **SIGINT.** Signals intelligence is gathered from the interception of transmissions, including communications intelligence, foreign instrumentation signals intelligence and electronic intelligence. SIGINT constitutes the greatest collection of intelligence because the majority of devices use electronic signals.
- **OSINT.** Open source intelligence is derived from sources which are publicly available. Such sources are news, commercial databases and social media. One type of OSINT are technical details about publicly accessible information like domain names or IP addresses and another type is published reports on cyber-security threats.
- **IMINT.** Imagery intelligence collects information from visual representations, like aerial photography and satellites or radars; typically, it is not a CTI source.
- **MASINT.** Measurement and signature intelligence is technical and scientific intelligence information that is derived from quantitative and qualitative analysis of information and includes intelligence on radio frequency, radar, acoustic, electromagnetic pulse, electro-optical, materials, chemical, biological, and laser intelligence, amongst others; likewise, it is not considered to be CTI source.
- **GEOINT.** Geospatial intelligence is intelligence concerning the human activity and it is collected by exploiting and analyzing imagery and geospatial information (like GPS data, maps, satellites) and any kind of information related to locations. IMINT is considered to be part of GEOINT and by this way GEOINT is also not a typical source of cyber threat intelligence.

Over the years various INTs have showed up, like cyber intelligence, financial intelligence, and technical intelligence, etc., but most of them are already covered by other intelligence collection methods.

2.3. CTI categories

CTI sources can be categorized into three different categories: *internal*, *community*, and *external*. In more details:

- **Internal.** This category encompasses any CTI that is collected from any organization. This may include reported information from security tools like *intrusion prevention systems* (IPS), firewalls, host security systems (anti-virus), etc. A significant source of threat intelligence information is derived from computer forensic analysis, which provides information about application settings, services being used, system events, etc., that could indicate adversarial behavior. Likewise, useful information is derived from honeypots and honeynets since they gather capture a wide range of attacks on a target system or network; they are classified as low or high interaction and internal or external (on the public internet) honeypots.
- **Community.** The community sources include CTI shared via a trusted relationship with multiple members having shared interests. This can be an informal group with member organizations that are in the same industry sector or that have other common interests. The *Information Sharing and Analysis Centers* (ISACs) are such an example [22]. They are non-profit organizations providing a central resource for gathering CTI and allowing two-way sharing between the private and the public sector.
- **External.** The external category contains CTI from sources outside an organization. There are three types of external sources.

- *Public sources* are external sources which are free of charges and available to anyone. Although public feeds are available at no cost, some issues can be occurred addressing volunteered data (such as the quality of these data).
- *Private sources* are not available without payment. An organization can subscribe to a threat feed from a vendor to receive a regularly updated CTI; these feeds may guarantee data quality based on a *service level agreement* (SLA). The majority of security products include some type of CTI update mechanism. For instance, the Emerging Threats ETPro Ruleset[1] offers subscription services for IP reputation and IDS rules.
- *Unindexed sources* are sites and forums from the deep or dark web. Companies having deep/dark web related CTI is a growing trend. In most cases, the CTI is gathered from chat-rooms and other forums with confined access that may be are neither easily accessible nor publicly available. In these sites, individuals usually exchange information that provides great security insights after being analyzed.

2.4. The CTI cycle

The CTI cycle, illustrated in Figure 2, is the formal process for generating and evaluating intelligence. The first step of this process is that of *CTI source identification* (it is also referred to as direction). It includes aspects pertaining to the identification of:

- threat information that needs to be collected as part of a monitoring strategy,
- sensors, feeds, and security repositories that generate threat information with the required frequency, precision and accuracy to support cyber-security decision-making,
- other threat information that is valuable but not necessarily updated on a regular basis and
- threat information suitable for sharing with external parties to raise cyber-security awareness.

The next step, *CTI gathering*, is the collection of the necessary data from the identified sources, along with the tools to build the capacity for collecting a wide variety of information, like tactical information (infrastructure, malware, and exploits) and strategic information (revealing attackers' goals). This step is performed in a continuous manner; it is not a one-time action (e.g., to gather information about domains requires a series of steps starting from the collection of an IP address). The main goal at this stage is to collect as much information as possible so as to allow correlations and further analysis.

[1] https://www.proofpoint.com/us/threat-insight/et-pro-ruleset

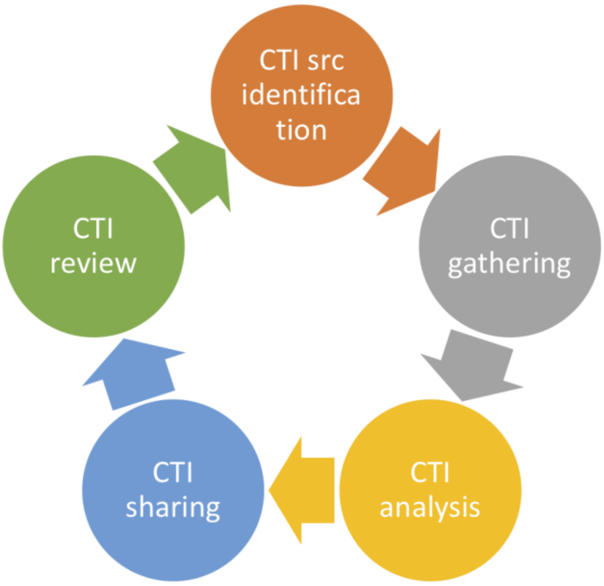


Figure 2. The intelligence life-cycle

The third step is *CTI analysis* and is built upon the information that has being collected; it includes both automated means of analysis (often referred to as CTI processing) by proper tools, and non-automated ones (also called CTI generation) that are carried out by human security experts and analysts. The use of tools to transform the collected raw data into a more usable and actionable format is necessary and this can also be considered as part of the CTI gathering step. Data related to cyber-threats are commonly processed in the following ways:

- **Filtering.** Not all data gathered are relevant and useful, and hence proper methods are used to filter out information that will not add any value in the analysis.
- **Normalization.** Refers to transforming the collected data into uniform formats so as to facilitate analysis, since CTI can be gathered in a variety of formats, like JSON, XML, CSV, and plain text.
- **Indexing.** Refers to making large volumes of gathered cyber-threat data searchable by the security analysts in an efficient manner.
- **Enrichment.** Includes the automatic characterization of the gathered data with additional relevant and contextual metadata (e.g., resolving domain names to IP addresses).
- **Prioritization.** The data gathered need to be ranked to drive the efforts of security analysts.

The above aspects are complemented by proper data visualization techniques for reducing the cognitive load of the security analysts. They subsequently assess the resulting data, possibly against other available data, in order to derive useful and accurate meanings as well as implications (and possibly predictions, e.g., by deriving emerging threat landscape trends) in cyber-security.

The fourth step is *CTI sharing* to the relevant stakeholders, i.e., the entities that can utilize the generated intelligence, in a form that they find to be appropriate, useful, and in many cases actionable. This makes sharing highly-dependent on the audience (e.g., tactical, operational, and strategic level — as illustrated in Table 1) and the goals.

CTI review, which is the last step in the above process (also referred to as CTI feedback), constitutes the key to the continuous improvement of the generated intelligence.

2.5. Uses of CTI

2.5.1. Impact on cyber-defense

The core idea of using threat intelligence is to earn an advantage over the adversary by detecting its presence, blocking or delaying its attacks, or degrading its infrastructure. A good way to describe the impact defenders are able to make was originally presented with the *pyramid of pain*, introduced by D. Bianco[2].

Every level in Figure 3 represents types of observables that defensive individuals can include in their blocking mechanisms. A hash value is not a very permanent indicator because attackers are in position to generate polymorphic code and change hashes by introducing null bytes in the code but some mitigating techniques might also help with that issue. In specific SSDEEP[3] and IMPHASH[4] are packing together similar files; SSDEEP brings a percentage of similarity based on the common bits of code, while IMPHASH hashes the table of these executable files and searches for similarities. Unfortunately, many security vendors do not adopt these methods and are slightly more difficult to use than the simple MD5.

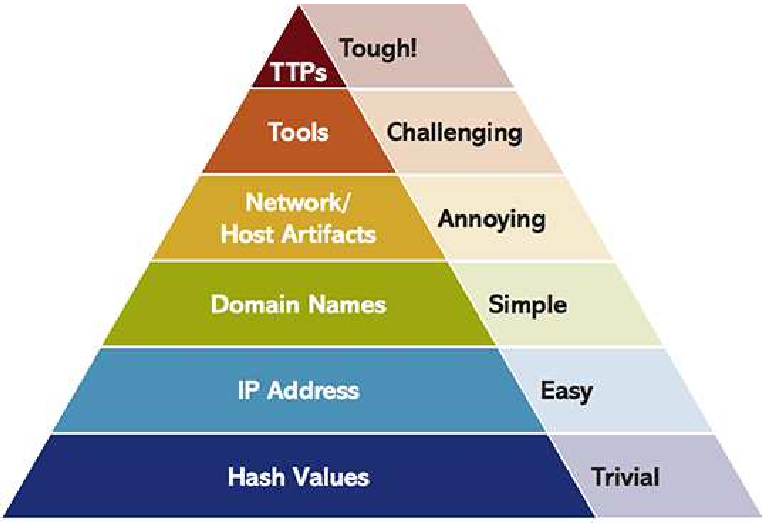


Figure 3. CTI impact on defensive strategies and associated gathering hardness

IoCs are the lowest data on the structure and thus the simplest to gather. Due to the fact that IoCs may occur in a variety of formats, they are defined as "a description of technical characteristics that identify a known threat, an attacker's methodology, or other evidence of compromise." IoCs focus on single pieces of information at the tail of the structure in Figure 3. On the other hand, behavior is more complicated to capture because it is difficult for the attackers to change their tactics and the techniques they use. As an example, a threat actor that specializes in attacking websites with SQL injection could find it hard to alter tactics and realize spear phishing with zero-day exploits. Hence, detecting and blocking a TTP, would force adversaries to change the way they actually do business and significantly increase their costs (e.g., due to

[2] http://detect-respond.blogspot.com/2013/03/the-pyramid-of-pain.html
[3] https://ssdeep-project.github.io/ssdeep/index.html
[4] https://www.fireeye.com/blog/threat-research/2014/01/tracking-malware-import-hashing.html

the need for changing tactical parts of their infrastructure and tools). The result is that when it comes to threat information, we prefer to have longer use out of the information toward the top of the structure. Our goal in intelligence analysis and incident response is to move higher up to the pyramid, which will make it more difficult for an adversary to evade.

2.5.2. Impact on decision-making

The impact that CTI has on the decision-making process lies at various levels.

- At **tactical level,** unverified IoCs generate false alerts and use flooding to obstruct analysts from identifying alerts that are linked to threats capable of causing significant damage.
- At **operational level** consuming for IR teams find relative information about threats, rebuild the attacks, and take actions to mitigate them.
- Finally**, at strategic level** CISOs and IT managers do not have the required information to set priorities or make budgeting and staffing decisions.

Table 1. Uses of CTI and importance towards helping defend against cyber-attacks [25]

	Tactical level	Operational level	Strategic level
IT roles	● Network operations center (NOC) ● Infrastructure operations ● Security operations center (SOC)	● Incident response (IR) team ● Security forensics ● Fraud detection	● Chief information security officer (CISO) ● IT management
Tasks	● Feed indicators to security products ● Patch vulnerable systems ● Monitor, escalate alerts (triage)	● Determine attacks' details ● Remediate ● Hunt for additional breaches	● Allocate resources ● Communicate with executive management
Problems	● Unverified indicators cause false positives ● Difficult to prioritize patches ● Too many alerts to investigate	● Time-consuming to reconstruct attacks from initial indicators ● Difficult to identify damage and additional breaches	● No clear priorities for investment ● Executives do not understand technical issues
CTI value	● Validate and prioritize indicators ● Prioritize patches ● Prioritize alerts	● Provide context to reconstruct attacks quickly ● Provide data to identify damage and other related breaches	● Provide priorities based on business risks and more likely attacks ● Thwart adversaries and threats

2.6. CTI sharing benefits

Threat information sharing provides access to threat information that in a different case might be unavailable to a small organization. There is a plethora of benefits in sharing CTI that include:

- **Increased situational awareness.** Using shared resources enhances security by leveraging, the capabilities of partners (knowledge, experience) in a proactive way.
- **Improved security posture.** It has become easier for organizations to identify affected systems, implement measures for protection, enhance the detection method in case an incident re-occurs and recover from incidents.
- **Knowledge maturing.** This enhanced process increases the value of information by enriching existing indicators and developing knowledge of actor TTPs that are associated with a specific incident, threat, or threat campaign.
- **Increased defensive agility.** Threat actors continuously adapt their TTPs and try to evade detection, security controls, and exploit new vulnerabilities. In order to reduce the probability of successful attacks, organizations are often informed about changing TTPs and usually detect and respond to threats rapidly.

2.7. CTI sharing challenges

As mentioned above sharing threat intelligence information has its benefits, but certain challenges still remain. A number of challenges that address the production and consumption of threat information are [52]:

- **Establishing trust.** Trust relationships create the basics for information sharing, but their maintenance requires significant effort.
- **Achieving interoperability and automation.** The use of standardized data formats is an important building block for interoperability because it allows organizations and repositories to easily exchange threat information standardized way. However, adopting specific formats may require significant time and resources.
- **Securing sensitive information.** Publishing sensitive information such as controlled unclassified data and personally identifiable information may result in violating sharing agreements, and loss of reputation, or even financial loss. Organizations should implement policies and technical controls to manage the risks of disclosure of sensitive data.
- **Enabling information consumption and publication.** All the organizations that are willing to publish and consume threat intelligence information should have the necessary tools and well-trained personnel to do so. Organizations unable to support automated indicator exchange formats for best practices should either use manual exchange or summary indicator information.

There are more challenges related to circumstances that need to be tackled [52] where an organization should use CTI. For example, how to access external sources and incorporate actionable CTI, how to estimate the quality of the received CTI and how to provide CTI (e.g., how to comply with policies or requirements pertaining to privacy and the limitation of attribution).

3. Description of methodology

This section presents the methodology used for providing recommendations on the CTI sharing aspects of the Cyber-Trust project by relying on a set of high-level requirements, as described in the *description of action* (DoA), and by considering the findings of desk research on the current situation on CTI sharing. This has taken into account:

- the availability of standards in 5. CTI formats and languages for reporting vulnerabilities, threats and other information gathered from various CTI sources presented in 4. CTI sources,
- their support from open source tools in 6. CTI platforms and tools based on which the Cyber-Trust platform will be developed, and
- the current market situation that is presented in detail in 7. Current market situation.

A distinction should be made between a sharing mechanism and a platform. While the former structures the encoding of information (e.g., by providing rules for XML tags to allow for automatic processing and possibly decision-making), the latter provides a tool allowing to efficiently share information. A number of requirements stemming from the above considerations are presented in the following sections.

3.1. Requirement 1: The sharing mechanism must allow CTI sharing between the Cyber-Trust platform and different stakeholders (service providers as well as CERTs, LEAs, CSIRTs, etc.)

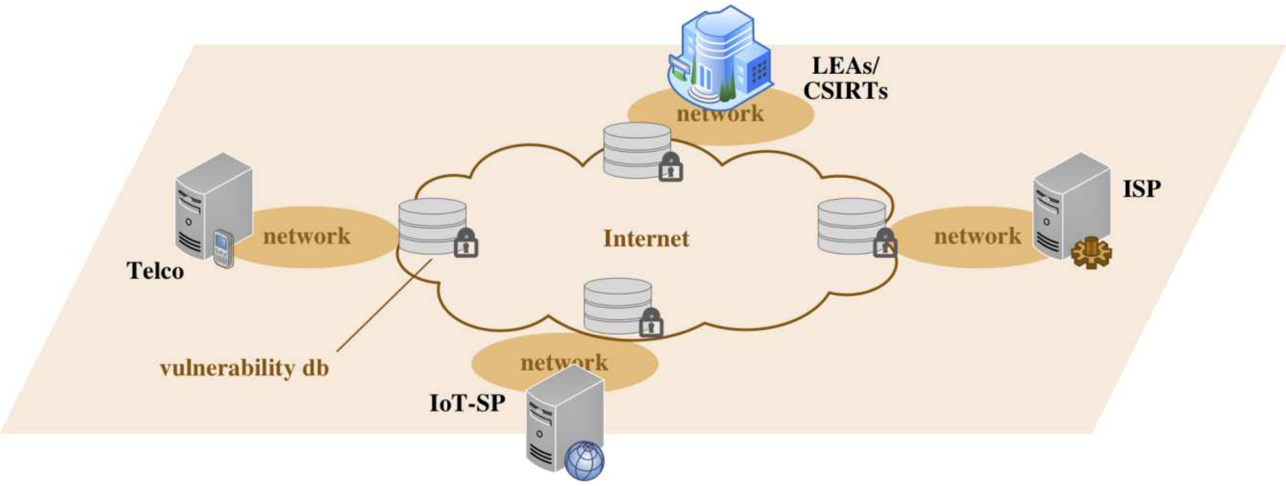


Figure 4. Sharing of CTI envisioned to be performed externally for increasing cyber-security awareness.

The common usage of a sharing mechanism with external entities could offer considerable advantages, such as automation in analyzing the data and in creating statistics. A number of existing initiatives for the exchange of cyber-threat related information between CERTs, LEAs, and CSIRTs should be considered to ensure high level of acceptance among the different stakeholders, as shown in Figure 4 (from the DoA). In this direction, ENISA has published a number of reports describing the current state of the art in this area (e.g., see [18,20]). Since there is no need to duplicate the analysis here, the outcomes stated therein are taken into account in 8. Recommendations, when making our recommendations.

3.2. Requirement 2: The sharing mechanism must allow CTI sharing between the Cyber-Trust platform and the end-users' devices, which requires industry's support

This is important for allowing the efficient exchange of information on ongoing cyber-attacks in order to implement the project's intelligent cyber-defense framework in Task T6.3, as shown in Figure 5 (from the DoA). This implies that the CTI sharing mechanism to be adopted should be widely accepted by the industry and already (heavily) used in existing products for, e.g., *security information and event management* (SIEM), intrusion detection or prevention systems. The landscape of standards and mechanisms available for sharing structured threat information was quite dynamic, until recently, with several proposals competing for adoption by the private sector. Nowadays, there exist mechanisms whose popularity are continuously growing and are turning into a de facto standard (as illustrated in 7. Current market situation and 8. Recommendations).

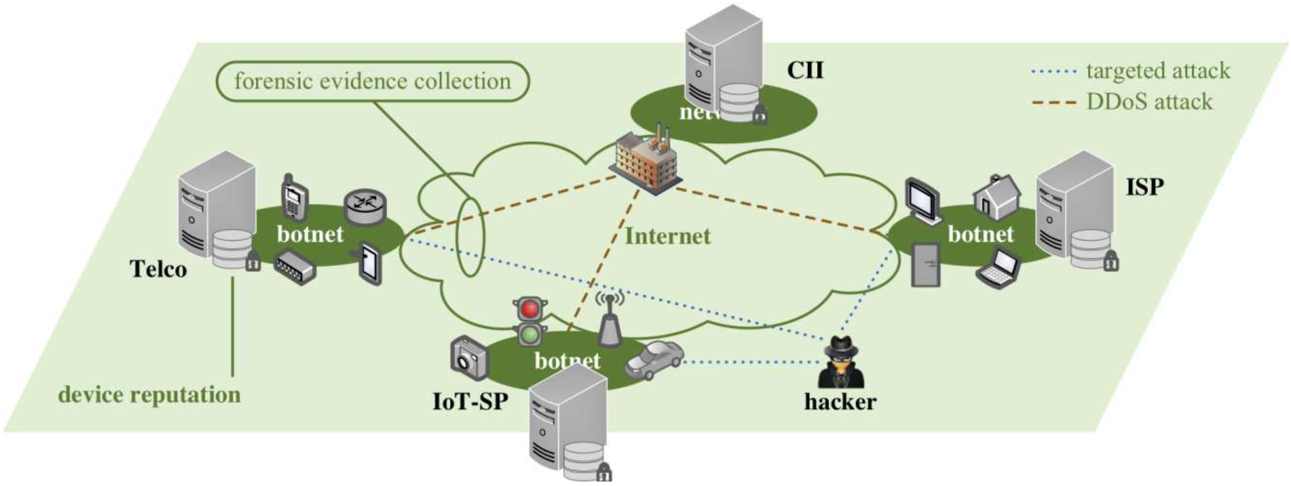


Figure 5. Sharing of CTI envisioned to be performed internally for allowing cyber-attack mitigation.

Meeting this requirement is a prerequisite for being able to support the needs of the use case scenarios and the end-user requirements, as they are currently being formed in the work of Tasks T2.2 and T2.3 respectively; the outcome of these efforts will be presented in the forthcoming Deliverables D2.3 and D2.4. Furthermore, adhering to industry standards is considered to be a competitive advantage for the subsequent exploitation of Cyber-Trust project's results in WP9.

3.3. Requirement 3: The sharing mechanism (and platform) should allow for a sufficient level of expressibility, flexibility, and scalability

The project's ambitious goal is to thwart a broad range of sophisticated cyber-attacks that are targeting IoT infrastructures, networks, and devices. Due to the high inherent complexity and heterogeneity of the IoT ecosystem, it is already evident that new forms of cyber-attacks, procedures, and threat actors' tactics are constantly emerging, which shape a broad threat landscape; the work that is being carried out in Task T2.1 is providing the current state-of-the-art in this area (included in Deliverable D2.1). However, the CTI sharing mechanisms of choice should be able to support the exchange of large volumes of information, not only for the attacks we are going to focus on in the course of the project (WP6), but also for future attacks as well.

These stress the need for structured representations of cyber-threat information meeting the above stated requirement.

3.4. Requirement 4. The sharing mechanism (and platform) should allow information to be both human and machine readable and facilitate automation

The exchange of cyber-threat related information in both a human-readable and machine-parsable form has a number of clear advantages. It allows the efficient and *automated* (software-based) data collection, parsing, filtering, categorization, and correlation, along with the subsequent thorough *analysis* by human security experts for the generation of cyber-threat intelligence that is usually quite hard (if possible) to be automated. Having the above two-stage process is essential in incident handling. In the context of the Cyber-Trust project this approach is embraced by having methods to be developed in WP5 and WP6 to:

- Leverage the benefits of machine-learning methods in the processing of very large amounts of raw data —either from external sources, e.g., from the surface/deep/dark web (T5.1), or internal, i.e., from an infrastructure's network (T6.3) and devices (T6.2).
- Drastically reduce the chance of overlooking critical security information, hence decreasing the false positive rate by using security experts to identify, highlight and analyze the data.

The intelligence analyst is then assisted by the envisaged graph-based visualization tool (T6.4) that will be used to assess the fidelity of cyber-threat information, as well as to discover, easily explore and understand complex information about the health status of an IoT network.

In addition to the above, machine readability is usually linked to the need for having CTI information that is *actionable*, and this in turn implies that there must be a clear indication of its origin for this to be achieved. As explained in 2.7. CTI sharing challenges, the receiver of CTI should be able to decide whether to *trust or not* the data based on their source. Therefore, the sharing mechanism should allow including details about the source of the information transmitted as well.

3.5. Requirement 5. The sharing mechanism should support some form of subscription method

Users of the Cyber-Trust platform should be able to receive the information relevant to their needs, their system's characteristics and configuration. In the DoA it was envisioned that devices' profiles will be developed; based on these profiles cyber-threat information sharing will be realized efficiently in a device-specific manner. The same holds for the sharing of CTI to external users, where they should be allowed to receive information only about the attacks of interest and on a specific subset of the data, e.g., observables and indicators without any details about the target or the threat actor.

3.6. Requirement 6. The sharing platform should be open source

This is a requirement stemming from the declared intention of Cyber-Trust to release key tools as open source software and make them available via GitHub.

4. CTI sources

In this section, we distinguish cyber threat intelligence sources into three categorizations: internal, community, and external. Internal sources are those within the organization, these sources have direct access to events and actions on systems and networks inside the external security perimeter of an organization. Community sources are those available from the community in which the organization exists, this can be partner organizations or open community-based sources; these typically consist of observed malicious sources or data, e.g., IP address, domain, URL, file names and hashes. External sources consist of data coming from external sources, which, although not immediately used by the organization, forms an integral part in understanding the wider threat landscape.

4.1. Internal sources of CTI

Identifying the internal sources of cyber threat information enables an organization to detect activity that indicates the presence of malicious actors or actions, which have a direct impact on the organization itself. Carrying out an inventory of internal and external sources will help identify any gaps on an organization's threat visibility, allowing the appropriate risk mitigation measures to be taken to an acceptable level.

Internal sources are able to produce large data-streams; in order to filter this kind of data and identify any significant events, automated tools (e.g., SIEM) should be employed. Analyzing and understanding the data is an ongoing process, and part of a continuous monitoring policy; this allows any rules to be fine-tuned to reduce false positives or make better use of the available sources.

The system's administrators and owners of the various sources should be consulted so that their knowledge, concerning what information is available and in what formats, can be leveraged to better understand potential threat. The ease of access and the format of the information can assist or inhibit the automated retrieval and processing of the data. The use of standard, structured and machine-readable formats will aid the processing, searching and any future retrieval of data.

Where indicators of compromise are detected, consideration should be given to the sharing of the data with the wider community. Organizations should ensure that any data shared is appropriately marked and sanitized of any confidential or personally identifiable information.

Table 2 lists selected internal information sources and example data that may be of use to security analysts:

Table 2. Selected internal information sources [52]

Source type	Source item	Information examples	Supported by Cyber-Trust

Source type	Source item	Information examples	Supported by Cyber-Trust
Host data sources	Operating system and application configuration settings, states, and logs	- Bound and established network connection and port - Process and thread - Registry setting - Configuration file entry - Software version and patch level information - Hardware information - User and group - File attribute (e.g., name, hash value, permissions, timestamp, size) - File access - System event (e.g., startup, shutdown, failures) - Command history	Yes
Host data sources	Antivirus products	- Hostname - IP address - MAC address - Malware name - Malware type (e.g., virus, spyware, remote access, etc.) - File name - File location (i.e., path) - File hash - Action taken (e.g., quarantine, clean, rename, delete)	Yes
Host data sources	Web browsers	Browser history and cache including: - Site visited - Object downloaded - Object uploaded - Browser extension installed or enabled - Cookies	No

Source type	Source item	Information examples	Supported by Cyber-Trust
Network data sources	Router, firewall, Wi-Fi, remote services (such as remote login or remote command execution), and Dynamic Host Configuration Protocol (DHCP) server logs	● Timestamp ● Source and destination IP address Domain name ● TCP/UDP port number ● Media Access Control (MAC) address hostname ● Action (deny/allow) ● Status code ● Other protocol information	Yes
Network data sources	Diagnostic and monitoring tools (network intrusion detection and prevention system, packet capture & protocol analysis)	● Timestamp ● IP address, port, and other protocol information ● Network flow data ● Packet payload ● Application-specific information ● Type of attack (e.g., SQL injection, buffer overflow) ● Targeted vulnerability ● Attack status (success/fail/blocked)	Yes
Other data sources	Security Information and Event Management (SIEM)	Summary reports synthesized from a variety of data sources, e.g. ● operating system ● application, and ● network logs	No
Other data sources	Email systems	Email messages: ● Email header content ○ Sender/recipient email address ○ Subject line ○ Routing information ● Attachments ● URLs ● Embedded graphics	No

Source type	Source item	Information examples	Supported by Cyber-Trust
Other data sources	Incident management/tracking system, and people from within the organization	● Analysis reports and observations regarding: ○ TTPs ○ Campaigns ○ Affiliations ○ Motives ○ Exploit code and tools ○ Response and mitigation strategies ○ Recommended courses of action ● User screen captures (e.g., error messages or dialog boxes)	Yes
Other data sources	Forensic toolkits and dynamic and/or virtual execution environments	● Malware samples ● System artifacts (network, file system, memory)	Yes

4.2. Community sources of CTI

Community-sourced observables and IoC feeds typically consist of observed malicious sources or data, e.g., IP address, domain, URL, file names and hashes. The principal use case is to use this information to create rule sets for firewalls, NIDSs, HIDSs, SIEM, etc., to block or alert on seeing the observable

Our desk research identified several sources of Open Source CTI along with the feeds they provide. The most prominent findings are listed below:

- abuse.ch
- Blocklist.de
- botvrij.eu
- CINS Army (Sentinel)
- Dshield (SANS)
- limo (Anomali)
- Malc0de
- PickUpSTIX
- Spamhaus
- TAXIIstand
- ÜberTAXII

4.2.1. abuse.ch

abuse.ch is a non-profit organization that provides several CTI feeds through the its web site [1]. These are provided through several projects:

- Ransomware tracker

- SSL blacklist
- URLHaus
- Snort/Suricata rules

We will briefly discuss these projects in the following sections.

4.2.1.1. Ransomware tracker

The ransomware tracker feeds provide a simple single column text file showing: Domain, URL or IP address as shown in Table 3. The tracker CSV feed is somewhat richer in content with multiple additional columns: Firstseen (UTC), Threat, Malware, Host, URL, Status, Registrar, IP address(es), ASN(s), Country.

Table 3. Ransomware tracker feeds

Feed / Blocklist	**Description**	**URL**
Domain blocklist	Domains associated with Ransomware	https://ransomwaretracker.abuse.ch/ downloads/RW_DOMBL.txt
URL blocklist	URL's associated with Ransomware	https://ransomwaretracker.abuse.ch/ downloads/RW_URLBL.txt
IP blocklist	IP addresses associated with ransomware payment sites	https://ransomwaretracker.abuse.ch/ downloads/RW_IPBL.txt
Tracker CSV feed	Combined CSV feed	https://ransomwaretracker.abuse.ch/ feeds/csv/

4.2.1.2. SSL blacklist (SSLBL)

SSL blacklist project (SSLBL) aims to provide a list of bad SSL certificates; bad is defined as an SSL certificate that is related to malware or botnet activities (e.g., botnet C&C traffic) [3]. SSBL can be downloaded in several formats (Text/CSV, STIX, DNS & Snort) as shown in Table 4 [3]. The feeds typically include either a SHA-1 fingerprint of an SSL certificate or the IP address of the server that used a bad certificate. The fingerprint feeds are provided by a text CSV file with multiple columns: Timestamp of Listing (UTC), SSL certificate SHA1 Fingerprint, Listing reason. The IP feeds are provided by a text CSV file with multiple columns: DstIP, DstPort, Reason.

Table 4. SSL blacklist feeds

Feed / Blocklist	**Description**	**URL**
Plain-Text SSL fingerprint blacklist (CSV)	Bad SSL Certificate SHA-1 fingerprint	https://sslbl.abuse.ch/ blacklist/sslblacklist.csv
Plain-Text SSL IP blacklist (CSV)	Bad SSL certificate IP Addresses	https://sslbl.abuse.ch/ blacklist/sslipblacklist.csv
Plain-Text SSL IP blacklist - aggressive (CSV)	Aggressive version, contains all IP's	https://sslbl.abuse.ch/ blacklist/sslipblacklist_aggressive.csv
Dyre Botnet C&C Bad SSL SHA-1 fingerprint	Dyre Botnet C&C Bad SSL SHA-1 fingerprint	https://sslbl.abuse.ch/ blacklist/dyre_sslblacklist.csv
Dyre botnet C&C SSL IP addresses	Dyre Botnet C&C SSL IP addresses	https://sslbl.abuse.ch/ blacklist/dyre_sslipblacklist.csv
Dyre botnet C&C SSL IP addresses aggressive	Dyre Botnet C&C SSL IP addresses aggressive	https://sslbl.abuse.ch/ blacklist/dyre_sslipblacklist_aggressive.csv
Dyre botnet C&C SSL	Common names (CN) of Dyre	https://sslbl.abuse.ch/

Feed / Blocklist	Description	URL
certificate common names	botnet SSL certificates	blacklist/dyre_commonnames.csv

4.2.1.3. URLHaus

This project provides a database of malware URLs that can be contributed to by community users, (Twitter account required), through an API [4]. The database dump file is a text CSV with the following fields: id, dateadded, url, url_status, threat,tags, urlhaus_link. The urlhaus_link URL will open the entry detail page on the URLhaus Database website. Quick reference of the feeds available and the respective direct URLs is shown at Table 5.

Table 5. URLHaus feeds

Feed / Blocklist	Description	URL
Database dump (CSV)	URLhaus database dump is a simple CSV feed that contains all malware URLs that are currently known to URLhaus	https://urlhaus.abuse.ch/downloads/csv/
Plain-Text URL list (URLs only)	A dump of all malware URLs known to URLhaus	https://urlhaus.abuse.ch/downloads/text/
Collected payloads (CSV)	All payloads collected by URLhaus, identified by a hash (MD5 / SHA256 hash)	https://urlhaus.abuse.ch/downloads/payloads/

4.2.1.4. Snort/Suricata rules

In addition to the text feeds, abuse.ch provides some lists as Snort/Suricata rules for use in the open source IDS/IPS systems. Quick reference of the feeds available and the respective direct URLs is shown at Table 6.

Table 6. Snort/Suricata rules

Feed / Ruleset	Description	URL
Suricata SSL fingerprint blacklist	Suricata SSL fingerprint blacklist	https://sslbl.abuse.ch/blacklist/sslipblacklist.rules
Suricata / Snort SSL IP blacklist	All hosts SSBL has seen in 30 days	https://sslbl.abuse.ch/blacklist/sslipblacklist_aggressive.rules
Dyre Botnet C&C Suricata SSL fingerprint blacklist	Dyre Botnet C&C Suricata SSL fingerprint blacklist	https://sslbl.abuse.ch/blacklist/dyre_sslblacklist.rules
Dyre Botnet C&C Snort SSL IP Blacklist	Dyre Botnet C&C Snort SSL IP blacklist	https://sslbl.abuse.ch/blacklist/dyre_sslipblacklist.rules
Dyre Botnet C&C Snort SSL IP blacklist - aggressive	Dyre Botnet C&C Snort SSL IP blacklist - aggressive	https://sslbl.abuse.ch/blacklist/dyre_sslipblacklist_aggressive.rules
URLhaus Snort / Suricata Ruleset	IDS Ruleset to identify network traffic towards known malware URLs	https://urlhaus.abuse.ch/downloads/ids/

4.2.2. Blocklist.de

Blocklist.de is a voluntary and free service that takes reports from over 4500 users that use fail2ban [24] and similar abuse blocking applications [8]. The lists can be accessed via an API, DNS Real-time Blackhole List (RBL) or via a simple text download. Files are a single column text file containing IP addresses. Quick reference of the Feeds available and the respective direct URLs is shown at Table 7.

Table 7. Blocklist.de feeds

Feed / Blocklist	Description	URL
all	All IP addresses that have attacked one of our customers/servers in the last 48 hours.	https://lists.blocklist.de/lists/all.txt
apache	All IP addresses which have been reported within the last 48 hours as having run attacks on the service Apache, Apache-DDOS, RFI-Attacks.	https://lists.blocklist.de/lists/apache.txt
bots	All IP addresses which have been reported within the last 48 hours as having run attacks on the RFI-Attacks, REG-Bots, IRC-Bots or BadBots (BadBots are bots that spam Forums or Wikis).	https://lists.blocklist.de/lists/bots.txt
bruteforcelogin	All IPs which attacks Joomla, Wordpress and other Web-Logins with Brute-Force Logins.	https://lists.blocklist.de/lists/bruteforcelogin.txt
ftp	All IP addresses which have been reported within the last 48 hours for attacks on the service FTP.	https://lists.blocklist.de/lists/ftp.txt
imap	All IP addresses which have been reported within the last 48 hours for attacks on the service imap, sasl, pop3, etc.	https://lists.blocklist.de/lists/imap.txt
ircbot	All IP addresses used by IRC-Bots	https://lists.blocklist.de/lists/ircbot.txt
mail	All IP addresses which have been reported within the last 48 hours as having run attacks on the service Mail, Postfix.	https://lists.blocklist.de/lists/mail.txt
sip	All IP addresses that tried to login in a SIP-, VOIP- or Asterisk-Server and are included in the IPs-List from http://www.infiltrated.net/ (Twitter).	https://lists.blocklist.de/lists/sip.txt
ssh	All IP addresses which have been reported within the last 48 hours as having run attacks on the service SSH.	https://lists.blocklist.de/lists/ssh.txt
strongips	All IPs which are older than 2 month and have more than 5.000 attacks.	https://lists.blocklist.de/lists/strongips.txt

4.2.3. botvrij.eu

This MISP powered site provides freely available IoCs, the information is gathered for various OSINT sources, including blog pages and PDF documents, and then consolidated into the different datasets [7]. The downloaded text files contain the observable value, (domain, IP, filename, etc.), followed by a comment describing the threat. The site also provides the CTI in MISP format; users and organizations who have the MISP platform installed can connect directly to botvrij.eu.
A quick reference of the feeds available and the respective direct URLs is shown at Table 8.

Table 8. Botvrij.eu feeds

Feed / Blocklist	**Description**	**URL**
ioclist.domain	Domains	http://www.botvrij.eu/data/ioclist.domain
ioclist.email-src	Email addresses (empty)	http://www.botvrij.eu/data/ioclist.email-src
ioclist.filename	Malicious filenames	http://www.botvrij.eu/data/ioclist.filename
ioclist.ip-dst	IP destinations	http://www.botvrij.eu/data/ioclist.ip-dst
ioclist.md5	File MD5	http://www.botvrij.eu/data/ioclist.md5
ioclist.regkey	Windows register key IoC's (empty)	http://www.botvrij.eu/data/ioclist.regkey
ioclist.sha1	File sha1	http://www.botvrij.eu/data/ioclist.sha1
ioclist.sha256	File sha256	http://www.botvrij.eu/data/ioclist.sha256
ioclist.url	URL	http://www.botvrij.eu/data/ioclist.url
MISP OSInt Feed	MISP format OSInt feed JSON files	http://www.botvrij.eu/data/feed-osint/

4.2.4. CINS Army by Sentinel

Sentinel Intrusion Prevention Systems provide an open source CTI list that is a subset of their commercial CINS Active Threat Intelligence ruleset [73] known as the *CINS Army List*. Data is collected from sentinel IPS devices deployed on customers' sites around the world, and contain addresses for which:

- The IP's recent rogue packet score factor is very poor
- The IP has tripped a designated number of trusted alerts

The list consists of IP addresses in a plain text file and is limited to 15,000 results, no further information or metadata is supplied on the IP addresses listed. The CINS Army List can also be downloaded as a tarball file containing the list in multiple formats: STIX v1.x, Snort rules and the text file that is available separately.

A quick reference of the feeds available and the respective direct URLs is shown at Table 9.

Table 9. CINS Army List feeds

Feed / Blocklist	Description	URL
CINS army list	Data collected from Sentinel IPS devices that have: a poor rogue packet score, tripped alerts	http://cinsscore.com/list/ci-badguys.txt
CINS army list	Download tarball containing CINS Army list in multiple formats: STIX v1.x, Snort, txt	https://cinsarmy.com/list-download/

4.2.5. Dshield by SANS

The Internet Storm Center (ISC), which is supported by the SANS Institute and manned by volunteers, provides the DShield API allowing the collection of data from intrusion log entries around the world and community contribution to the log sources [67]. The ISC provides a threat feed web portal that details the threat feed data. The threat intelligence feeds are available in multiple formats, (XML, JSON and text), via a REST API [15], and can be downloaded as the tab-delimited plain text data files described in Table 10.

Table 10. DShield text feeds

Feed	Description	URL
Recommended block list	This list summarizes the top 20 attacking class C (/24) subnets over the last three days	http://feeds.dshield.org/block.txt
Current most active port scanning IPs	Top 10 most wanted, IP and resolved domain if available	http://feeds.dshield.org/top10-2.txt
Current most scanned ports	DShield.org top 10 target ports	http://feeds.dshield.org/topports.txt
Suspicious domain list (high)	This list consists of high level sensitivity website URLs	http://feeds.dshield.org/ suspiciousdomains_High.txt
Suspicious domain list (medium)	This list consists of medium level sensitivity website URLs	http://feeds.dshield.org/ suspiciousdomains_Medium.txt
Suspicious domain list (low)	This list consists of low level sensitivity website URLs	http://feeds.dshield.org/ suspiciousdomains_Low.txt

4.2.6. Limo by Anomali

Anomali provides commercial applications and services [6] alongside its STAXX open source platform and Limo TAXII CTI feed [5]. The Limo free threat feed supports TAXII v2.0 and provides the threat intelligence in STIX v2.0 JSON format. Several collections are available as shown in Table 11 below with the TAXII endpoint URLs.

Table 11. Anomali Limo TAXII Collections

Collection	Description	URL
Abuse.ch ransomware domains	Ransomware URL blocklist, IP addresses associated with	https://limo.anomali.com/api/v1/taxii2/feeds/collections/136/

	ransomware payment sites	
Abuse.ch ransomware IPs	Ransomware IP blocklist, IP addresses associated with ransomware payment sites	https://limo.anomali.com/api/v1/taxii2/feeds/collections/135/
Anomali weekly threat briefing	Anomali weekly threat briefing	https://limo.anomali.com/api/v1/taxii2/search_filters/collections/1/
Blutmagie TOR nodes	TOR exit nodes	https://limo.anomali.com/api/v1/taxii2/feeds/collections/209/
CyberCrime	CyberCrime feed	https://limo.anomali.com/api/v1/taxii2/feeds/collections/41/
DShield scanning IPs	Scanner IP addresses	https://limo.anomali.com/api/v1/taxii2/feeds/collections/150/
Emerging threats - compromised	Compromised servers	https://limo.anomali.com/api/v1/taxii2/feeds/collections/68/
Emerging threats C&C Server	C2 servers	https://limo.anomali.com/api/v1/taxii2/feeds/collections/31/
Lehigh malware domains	C2 servers	https://limo.anomali.com/api/v1/taxii2/feeds/collections/33/
Malware domain list - hotlist	Malware domain list - Hotlist domains	https://limo.anomali.com/api/v1/taxii2/feeds/collections/200/
Phish tank	Phishing URLs	https://limo.anomali.com/api/v1/taxii2/feeds/collections/107/

4.2.7. Malc0de

The Malc0de database provides a list of domains and IP addresses that have been identified as having distributing malware during the past 30 days. The lists are updated daily and can be accessed and searched via the Malc0de database web portal [37] or downloaded in several formats. An RSS feed, with URL, IP address, country ASN, and MD5 hash is also available. The URL points back to the entry detail page on the Malc0de database web portal. IP blacklist is a simple text file containing malicious IP addresses. Malc0de offers also DNS zone files in BIND and Windows formats that can be loaded into DNS servers so that DNS requests to Malware domains return 127.0.0.1.
A quick reference of the feeds available and the respective direct URLs is shown at Table 12.

Table 12. Malc0de database feeds

Feed	Description	URL
Database RSS feed	Sources of malicious executable, URL, IP, country, ASN, Hash	http://malc0de.com/rss
IP Blacklist	IP blacklist of hosts serving malicious executables	http://malc0de.com/bl/IP_Blacklist.txt
BIND format	BIND DNS blacklist zone file	http://malc0de.com/bl/ZONES
Windows Format	Windows DNS blacklist zone file	http://malc0de.com/bl/BOOT

4.2.8. PickUpSTIX by NC4 / Soltra

PickUpSTIX is an open source TAXII CTI feed provided by NC4 / Soltra [45]. These collections are provided via a TAXII v1.x interface and delivered in STIX v1.x XML format.
A quick reference of the feeds available and the respective direct URLs is shown at Table 13.

Table 13. PickUpSTIX TAXII Collections

Collection	Description	URL
cybercrime-tracker.net	Cyber crime tracker botnet controllers	https://pickupstix.io/taxii-discovery-service/service/cybercrime-tracker.net
Default	Combined feed ransomware and Malc0de	https://pickupstix.io/taxii-discovery-service/service/Default
malc0de_org	Hosts serving malicious executables	https://pickupstix.io/taxii-discovery-service/service/malc0de_org
ransomwaretracker_abuse_ch	Ransomware	https://pickupstix.io/taxii-discovery-service/service/ransomwaretracker_abuse_ch

4.2.9. Don't Route Or Peer Lists by Spamhaus

The Spamhaus Don't Route Or Peer Lists (DROP) are probably one of the oldest and best-known CTI sources. They consist of netblocks that are hijacked or leased by professional spam or cyber-crime operations [74]. The downloadable text drop lists are a subset of the full Spamhaus Block List (SBL) aimed at network filtering devices.

Table 14 shows the text DROP lists available along with a brief description and the URL. These lists are open and free to use.

Table 14. Spamhaus DROP Lists

List	Description	Download URL
Don't Route Or Peer List (DROP)	The DROP list will not include any IP address space under the control of any legitimate network - even if being used by "the spammers from hell"	https://www.spamhaus.org/drop/drop.txt
Extended DROP list (EDROP)	EDROP is an extension of the DROP list that includes suballocated netblocks controlled by spammers or cyber criminals. EDROP is meant to be used in addition to the direct allocations on the DROP list.	https://www.spamhaus.org/drop/edrop.txt
IPv6 DROP list (DROPv6)	The DROPv6 list includes IPv6 ranges allocated to spammers or cyber criminals.	https://www.spamhaus.org/drop/dropv6.txt

ASN DROP list (ASN-DROP)	ASN-DROP contains a list of Autonomous System Numbers controlled by spammers or cyber criminals, as well as "hijacked" ASNs. ASN-DROP can be used to filter BGP routes which are being used for malicious purposes.	https://www.spamhaus.org/drop/asndrop.txt

4.2.10. TAXIIstand by EclecticIQ

The TAXIIstand Open TAXII service is provided by EclecticIQ [16] to promote the use of the STIX v1.x and TAXII v1.x open standards [17]. They also provide the open source OpenTAXII server and Cabby TAXII client. Only a single feed is currently available via TAXII v1.x, supplied in STIX v1.x XML format. Quick reference of the collections available and the respective direct URLs is shown at Table 15.

Table 15. TAXIIstand Collection

Collection	Description	URL
VxVault	Malware URI List	https://open.taxiistand.com/services/discovery

4.2.11. ÜberTAXII by Kingfisher Operations

The ÜberTAXII service is provided by Kingfisher Operations [36] who assists organizations wishing to engage with the STIX and TAXII v2.0 standards.
A quick collection of the feeds available and the respective direct URLs is shown in Table 16.

Table 16. ÜberTAXII Collections

Collection	Description	URL
CIRCL	Computer Incident Response Center Luxembourg (CIRCL) feed	https://ubertaxii.com/taxii/circl/
AIS	Department of homeland security AIS feed	https://ubertaxii.com/taxii/ais/
Perch	Perch security	https://ubertaxii.com/taxii/perch/
MITRE	MITRE ATT&CK data	https://ubertaxii.com/taxii/mitre/

4.3. External sources of CTI

This section outlines sources of threat intelligence external to the organization and the community, that contribute to the building of the threat landscape. This CTI is at a higher level, principally used by analysts to assess the threat landscape and determine the risks presented to the organizations, systems and networks. main classes of sources include:

- News feeds on articles covering ongoing threats.
- Vulnerability alerts and advisories.
- Search automation using search technologies to find vulnerable systems.
- Information on malware.

- Intelligence available directly from the dark web.

In the next subsections we will analyze the aforementioned categories.

4.3.1. News feeds

News items provide a wider and general awareness of current threats and events that may increase or reduce the risk to a given country, sector or organization. News feeds are provided in several formats that include: email, web sites and RSS feeds provided by industry news and magazine channels, national and regional CERTs.

Regular cyber-security news updates increase awareness to threats, they can also provide warning of current outbreaks or specific threats and on some cases the necessary actions and remediations to take. RSS feeds can contribute to Information overload and distraction from non-applicable news items is always a danger. The RSS feeds do not have a standardized or recognized taxonomy to aid automated filtering. Some examples of RSS feeds are given in Table 17.

Table 17. Example news feeds

Source	URL	RSS
CERT-EU	https://cert.europa.eu/cert/filteredition/en/CERT-LatestNews.html	https://cert.europa.eu/rss?type=category&id=CERT-LatestNews&language=en&duplicates=false
UK National Cyber Security Centre weekly report	https://www.ncsc.gov.uk/index/report	https://www.ncsc.gov.uk/feeds/reports.xml
nakedSecurity by Sophos	https://nakedsecurity.sophos.com/	https://nakedsecurity.sophos.com/feed
Security Boulevard	https://securityboulevard.com/	https://securityboulevard.com/feed/
ThreatPost	https://threatpost.com/	https://threatpost.com/rss-feeds/
Schneier on Security	https://www.schneier.com/	https://www.schneier.com/blog/atom.xml
Krebs on security	https://krebsonsecurity.com/	https://krebsonsecurity.com/feed/
DarkReading	https://www.darkreading.com/	https://www.darkreading.com/rss_feeds.asp
SecLists.Org security mailing list archive	http://seclists.org/	Multiple RSS feeds listed
SANS Institute - Newsbites	https://www.sans.org/newsletters/newsbites/	https://www.sans.org/newsletters/newsbites/rss/

News feeds are a valuable source of CTI, feeds should be chosen for best applicability and quality to avoid information overload.

4.3.2. Vulnerability advisories and alerts

Many vendors produce lists of vulnerabilities and exposures. The *Common Vulnerabilities and Exposures* (CVE) is the principal registry for vulnerabilities and other related information. Since the launch of CVE in 1999 over 100,000 entries for vulnerabilities and exposures have been created. Each CVE contains: (a) a

CVE ID number, e.g., CVE-2018-1234, (b) a brief description, and references to sources, either vendor specific or third-party vulnerability reports. A list of all CVE entries can be downloaded in various formats: CSV, HTML, Text, 'cve_1.0.xsd' or CVRF [40]. The recommended source is the National Vulnerability Database (NVD) [53] which provides a search facility for CVE and offers downloads in several formats (JSON, XML) and filtered data sets: year/month, latest, latest updated.

A list of vulnerability and security advisory sources, in addition to the CVE and NVD, are shown in Table 18; such sources typically offer various download formats and RSS or CVRF feeds.

Table 18. Vulnerability advisories and security alerts

Source	Description	URL
CERT	Vulnerability notes database	https://www.kb.cert.org/vuls
Microsoft	Security portal	https://portal.msrc.microsoft.com/en-us/security-guidance
Cisco	Cisco security advisories and alerts	https://tools.cisco.com/security/center/publicationListing.x
Oracle	Oracle security and patch update advisories	https://www.oracle.com/technetwork/topics/security/alerts-086861.html
Red Hat	Red Hat security advisories	https://access.redhat.com/security/vulnerabilities
SecurityFocus	Symantec SecurityFocus vulnerabilities, inc. Bugtraq	https://www.securityfocus.com/vulnerabilities

The CVE approach provides a standard format for vulnerabilities and security advisories that has a broad acceptance in the industry. The NVD CVE data also includes references to the type of weakness that is exploited via CWE [48] and the platform software and configuration effected through the CCE details list [51] which at this time does not appear to be regularly updated or well supported. In comparison the vendor-based sources analyzed are highly descriptive of which platform and software version is impacted.

The lack of a well-supported common taxonomy to describe the applicability of the vulnerability presents additional effort for security analysts and administrators who need to determine the applicability of the vulnerability and take appropriate action.

Many OSs and software packages provide their own in-house vulnerability feeds in their own format, RSS and CVRF. These are often in parallel to the CVE mechanism and are also used to distribute non-critical software update advisories. CERT advisories are usually distributed via news feeds or Twitter.

The CVE system is used globally and represents a success in CTI standards and procedures, although subject to enhancement by various parties the core reporting, disclosure and distribution of advisories is supported throughout the industry.

4.3.3. Search automation

Search engines are not a typical threat feed, engineered searches can be automated and used to provide lists of host IPs or URLs that are at risk to known vulnerabilities or that may already be compromised through detection of backdoor ports.

4.3.3.1. Google

Web search engines continuously crawl the web and the results are stored in data centers. This can often reveal a great deal of information on the security offered by a website, since specific search parameters passed to Google can leak sensitive information, locate private files, allow access to directories supposed to be accessible only to their owners, passwords, configurations, etc. These are known as Google Dorks, a comprehensive list of these can be found at the Google Hacking Database [61].

4.3.3.2. Shodan

Shodan is a search engine that crawls the entire Internet at least once a month, with on-demand scanning capabilities [72] to identify open and vulnerable hosts on the internet. The Shodan scans of the Internet are wide ranging, this can contribute to the detection of high risk or compromised hosts and contribute to the portfolio of services. The rate of scanning and therefore the speed of detection may be too slow to detect some threats or vulnerabilities in an effective time frame.

Data are available via a limited Web interface, CLI or API, that provides comprehensive access to data and can be automated with new discoveries streamed as they are discovered. The Shodan developer portal [71] has various levels of paid access for commercial use, based on the number of on-demand scans and results downloaded per month. Enterprise bulk data access is also available for downloading all Shodan data.

4.3.3.3. Shodan alternatives

Two other Shodan-like sites, which were identified during our research, are provided in
Table 19.

Table 19. Vulnerability advisories and security alerts

Site	Description	URL
Censys	Shodan type search and analyze engine	https://censys.io/
Ichidan	Dark web scanner, (appears to be down)	http://ichidanv34wrx7m7.onion

If the target network or devices are known, direct scanning with tools may return more comprehensive and in-depth analysis; however, such actions require a higher skill set. Such example tools include nmap or IoTSeeker, [64], and vulnerability scanners like Nessus [76] or OpenVAS [27].

4.3.4. Malware threat intelligence

Detailed intelligence on malware threats, TTP's, methods of infection, files, registry locations and other observables provides analysts the means to detect malware infections and protect target systems and networks from the spread of any infection. Two types of source were identified for malware CTI: antivirus suppliers (
Table 20) and hosted sandbox services (Table 21).

Table 20. Anti-virus supplier threat intelligence

Source	URL	RSS feed

Source	URL	RSS feed
Sophos Labs, viruses and spyware	https://www.sophos.com/en-us/threat-center/threat-analyses/viruses-and-spyware.aspx	https://www.sophos.com/en-us/rss/threats/latest-viruses.xml
McAfee threat landscape dashboard	https://www.mcafee.com/enterprise/en-gb/threat-center/threat-landscape-dashboard.html	n/a
Trend Micro, threat encyclopaedia and malware blog	https://www.trendmicro.com/vinfo/gb/threat-encyclopedia/	http://feeds.trendmicro.com/Anti-MalwareBlog
Symantec Security Response RSS feeds - Threats, risks, virus definition stats, vulnerabilities	http://securityresponse.symantec.com/avcenter/cgi-bin/syndicate.cgi	http://www.symantec.com/xml/rss/listings.jsp?lid=latestthreats30days
Malwarebytes Labs, threat analysis	https://blog.malwarebytes.com/threats/	https://blog.malwarebytes.com/feed/

Table 21. Hosted sandbox services

Source	URL
VxStream Sandbox	https://www.hybrid-analysis.com/
VirusTotal	https://www.virustotal.com/
Cuckoo Sandbox	https://cuckoosandbox.org/

Both the antivirus suppliers and hosted sandbox services can provide detailed analysis of malware behavior; detail pages are accessed via a web portal. Several of the AV suppliers offer RSS feeds, most of which are blog feeds and not based on a specific CTI format like MAEC. The sandbox sites offer limited free access via a web portal they provide detailed analysis of malware for a given URL, file upload or search. Limited download capability exists with some support for standard formats such as MAEC.

No useful open source or community automated CTI feeds were identified. Most AV suppliers offer these as part of their paid-for service offering. Retrieval of detailed analysis is predominantly a manual process.

4.3.5. Dark web

TOR [77] was created to enable privacy and anonymity on the internet, allowing citizens and news reporters in repressed or censored regimes to access otherwise blocked destinations safely without being tracked or persecuted. Despite its pure intentions this functionality also allows members of organized crime and other nefarious activities such as cyber criminals to use the same system to remain anonymous. Based on these activities the onion web has become known as the dark web: a web of sites where malware, botnets, hacking services, etc., can be bought and sold, methods and activities discussed. These sources can be valuable to skilled security analysts at the leading edge of cyber-security.

The dark web search concentrated on finding intelligence, tools and services that are not available on the normal web. The dark web does not possess the same search facilities that are commonplace on the surface web, to carry out the search for tools several search and link list sites were used.

As noted, the dark web due to the anonymity it offers, is used for criminal and illegal activities which is useful for locating useful CTI sources. Search / link resources are Illustrated in

Table 22.

Table 22. Dark web search resources

Site	Comments	URL
TechIncidents	Clear-web site containing dark web links, better than most of the 'wiki' link sites on the dark web	https://techincidents.com/dark-web-websites/
DuckDuckGo	Mainly clear web results	http://3g2upl4pq6kufc4m.onion/
Torch	Illegal finance and drugs	http://xmh57jrzrnw6insl.onion/
DeepDotWeb	News and market site, no useful links	http://deepdot35wvmeyd5.onion/
TorLinks	Hacking links tried, problems loading	http://torlinkbgs6aabns.onion/
AHMIA	Several interesting results, reviewed below	http://msydqstlz2kzerdg.onion/
Not Evil	Several interesting results, reviewed below	http://hss3uro2h4mxog3j.onion/
TheHiddenWiki	No useful information	http://zqktlwi4fecvo6ri.onion/
FreshOnions	Very long site list, poor metadata.	http://zlal32teyptf4tvi.onion/
HD Wiki	Illegal finance, drugs, market, message, social network links, IRC/Jabber, etc.	http://hdwikicorldcisiy.onion/
Grams	Problems loading	http://grams7enufi7jmdl.onion/
DeepWebSiteLinks	Circular links	http://deepweb2teloq5cl.onion/
SecMail	Anonymous email service	http://secmailw453j7piv.onion/

The reliability of search and listing sites was, at best, poor, with over 75% of links followed timing out. An anonymous email address was configured for use where sites required sign-up to preserve the anonymity of the researcher.

Table 23 presents a sample of the dark web sites visited that appeared most likely to have some useful tools or content.

Table 23. Dark web Sites

Site	Comments	URL
CodeGreen	Hacktivist website with Tools section, all available for Kali	http://pyl7a4ccwgpxm6rd.onion
X4Priv8	Links for tools, all require Fas.li add-on	http://opqy6r3vy6qft26e.onion/

Site	Comments	URL
	installed. Not followed	
Hidden Clubs	Request to join message boards, malware club revealed Mirai source code on GitHub.	http://x7giprgefwfvkeep.onion/
BlackHost	Tools, pastebin, email, mailer, webproxy	http://blackhost5xlrhev.onion/
Brotherhood Hackers	Various exploit tools, penbox tools, comprehensive toolset	http://wqekut2pocn45hwp.onion/
Digital Gangster	Numerous tools and services	http://psj55rofc5dcsod7.onion/ digitalgangsters/index.html
0day Database	List of vulnerabilities and exploits, shellcode, has custom Kali but mainly CVE's	http://odaydbcthpmtnqym.onion/
IntelExchange	Message board, Q&A on dark web	http://rrcc5uu3dkhlvdwo.onion
Hacker 4 Hire	One of many hackers for hire willing to do anything for money	hacker4hhjvre2qj.onion

Due to the nature of dark web, websites are not indexed and continuously change to new URLs. Many of the bulletin boards that suggest more lucrative CTI are by invitation only, and these were not examined further; such forums usually require that users gain some credibility before being able to access more privileged information and interact with other users. A number of websites containing information about tools used by such communities (and possibly exploits that are of interest to the Cyber-Trust project) were identified during this preliminary research, an initial selection of which includes:

- **Brotherhood Hackers**. A useful looking toolset called penbox found with tools for all stages of penetration testing. All the tools checked in the toolset were found to be available on the internet.
- **CodeGreen**. A hacktivist site with common surface web tools, NMap, W3af, and others.
- **Digital Gangster.** An interesting toolset, but further analysis of the python and PHP code revealed that the tools were commonly available on Github.
- **Hidden Clubs**. This is an invite-only message board. Access was granted on request, however the content was of little interest (e.g., only a GitHub link for the Mirai Botnet source code); a simple search of GitHub is considerably quicker and more lucrative.

Undoubtedly, the dark web is a really good resource for getting information about the tactics, techniques and procedures employed by cyber-threat actors, and for enriching the project’s vulnerability DB (VDB) with zero-day vulnerabilities and related CTI.

5. CTI formats and languages

The complexity of the IoT ecosystem leads to a cyber-threat landscape of growing sophistication and complexity, where cyber-security incidents occur with increasing frequency. This fact necessitates efficient and automated tools for analyzing and sharing heterogeneous CTI related to the present systems' configurations, attackers' threats and tactics, indicators of ongoing incidents, etc., in order to build proper and effective defensive capabilities. Given the numerous architectures, products and systems being used as sources of data for information sharing systems (as it is already presented in 2. CTI sharing overview), standardized and structured CTI representations are required to allow a satisfying level of interoperability across the various stakeholders.

As highlighted in several works (*see* e.g., [30, 66]) considerable efforts have been put during the last decade to standardize the data formats and exchange protocols related to CTI; the initiative led by MITRE, referred to as *making security measurable* (MSM) (https://msm.mitre.org), constitutes the most prominent such effort along with the more recent initiatives of ENISA towards improving cyber-threat information sharing among the CERTs, CSIRTs, LEAs, and other relevant stakeholders [18, 20, 21]. An overview of existing efforts is given in Figure 6, where standards are classified into different areas, that also depicts the variability in the areas covered by standards.

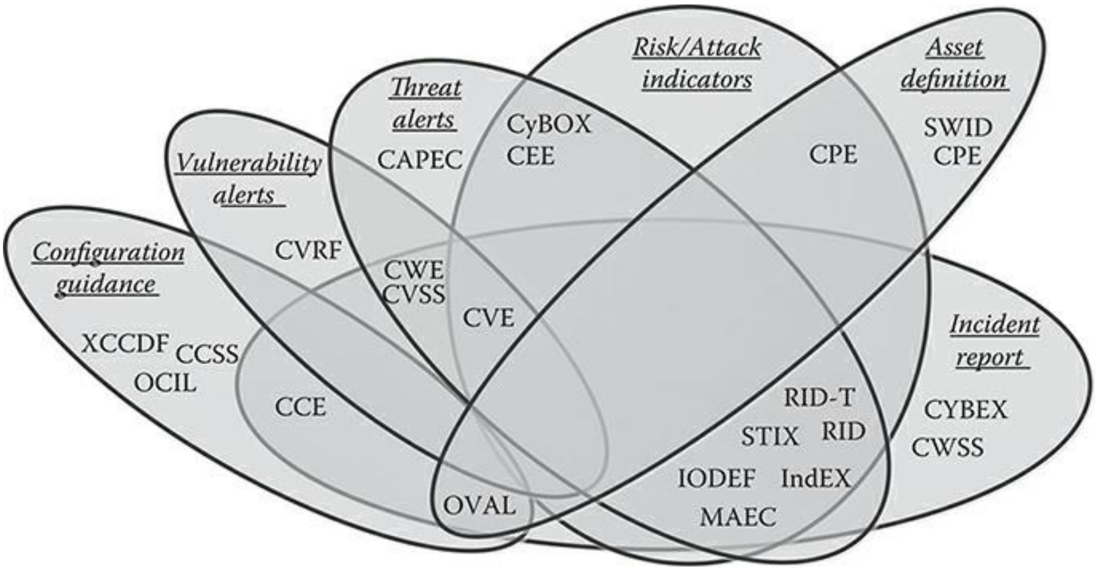


Figure 6. Areas covered by the different existing standards [30]

Some of the standards define the way CTI should be described; they are mostly based on the exchange of IoCs. After IoCs have been identified, they can be shared for detecting future attack attempts. Initiatives standardizing the format of the IoC descriptions shared allow their efficient and automated processing (*see* also [19]). In the following subsections, we briefly describe prominent initiatives, including others not mentioned above, like IDMEF that precedes IODEF and is still used by many tools, OPENC2 that is a cyber-defense coordination framework, and VERIS that focuses on strategic and risk-based information.

5.1. STIX (Structured Threat Information Expression)

The *Structured Threat Information Expression* (STIX) v2 is based on, but, supersedes STIX v1.x with the principal changes being the change in serialization from XML, to JSON, designed to make the protocol simpler for programmers and more lightweight. The structure is now flat rather than nested, with STIX Domain objects, (SDO), defined at the top level of the document to simplify parsing and storage. The relationship between the SDO's is accommodated by the introduction of a STIX Relationship Object (SRO) [54]. CybOX is now integrated into STIX; its objects have become "Cyber Observable" objects in STIX and are shared with MAEC, deducing complexity.

5.1.1. Use case

STIX v2.0 is designed to enable the sharing of CTI that should be machine readable and consistent. This in turn, allows the better understanding of attacks and enable a rapid and effective response and mitigation.

5.1.2. Data model

STIX uses a graph-based data model; the root node is always *bundle*, with the top-level objects being classified as *STIX Domain Objects* (SDO), *STIX Relationship Objects* (SRO), and marking definitions. The STIX observable_objects shown in Figure 7 represent cyber-observable data types, like IP address, file, etc. They are also used in STIX patterns in the indicator SDO, to match patterns of data, which may be described by the use of the observed-data object structure.

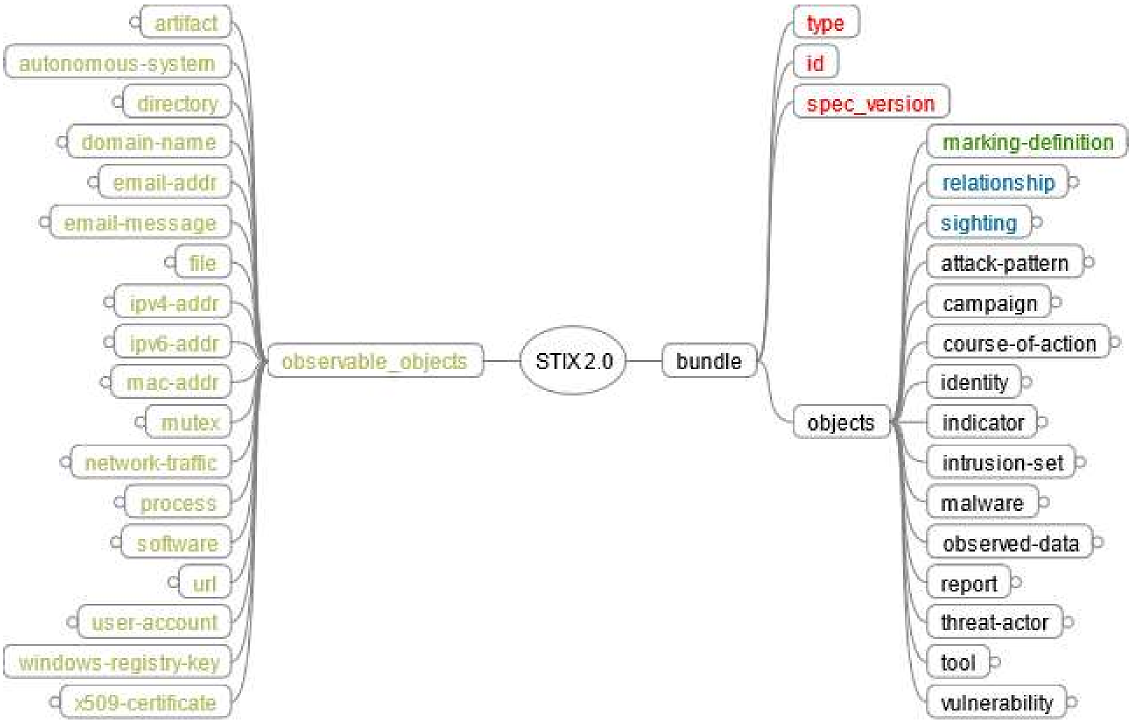


Figure 7. STIX 2.0 data model

5.1.2.1. STIX domain objects

The twelve STIX domain objects, shown in Figure 7, are used to represent different forms of CTI, they are represented as shown in Table 24 [54, 56].

Table 24. STIX domain objects

	SDO	**Description**
	Attack pattern	TTP that describe ways that adversaries attempt to compromise targets
	Campaign	A grouping of adversarial behaviors that describes a set of malicious activities or attacks
	Course of action	Stub: action taken either to prevent an attack or to respond to an attack
	Identity	Represent actual individuals, organizations, or groups
	Indicator	Contains a pattern that can be used to detect suspicious or malicious cyber activity, may use observable-object paths to describe data.
	Intrusion set	A grouped set of adversarial behaviors and resources with common properties that is believed to be orchestrated by a single organization.
	Malware	Malware description only, intended to represent malware, currently no link or reference to MAEC.
	Observed data	Information that was observed on systems and networks using the Cyber Observable specification, e.g., IP address, file, URL, etc.
	Report	Collections of threat intelligence focused on one or more topics, such as a description of a threat actor, malware, or attack technique, including context and related details.
	Threat actor	Actual individuals, groups, or organizations believed to be operating with malicious intent.
	Tool	Legitimate software that can be used by threat actors to perform attacks.
	Vulnerability	A flaw in software that can be directly used by a hacker to gain access to a system or network, e.g., CVE

5.1.2.2. STIX relationship objects

The two STIX relationship objects, shown in Figure 7, describe the relationships between the SDO's:

Table 25. STIX relationship objects

	SRO	Description
Relationship	Relationship	Describes the relationship between two SDO’s
Sighting	Sighting	Describes the relationship between SDO’s where CTI was seen or sighted

The STIX 2.0 architecture diagram shown in Figure 8 shows how the SDO’s are related to each other, e.g., a threat actor uses an attack-pattern, that uses malware, in attack. This enables SDO’s to be linked to fully describe a single attack or the campaigns orchestrated by an adversarial group.

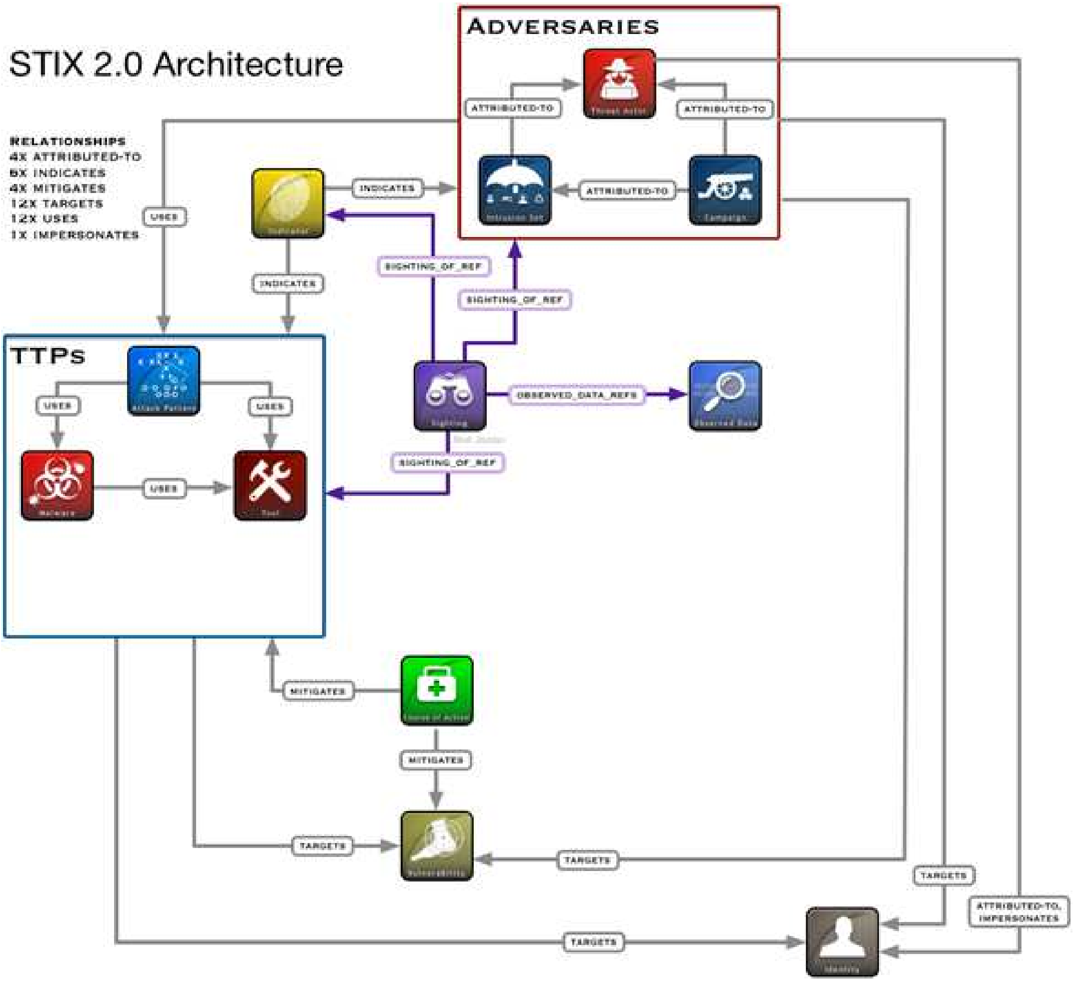


Figure 8. STIX 2.0 architecture [35]

5.1.2.3. STIX Patterns

STIX patterns are used by the indicator SDO’s to determine whether a given expression is true or false. Expressions consist of operators (comparison, set, and observation), object paths (cyber observable objects such as network-traffic:dst_port), observation expressions (e.g., [ipv4-addr:value = '198.51.100.1/32' OR ipv4-addr:value = '203.0.113.33/32']) as well as qualifiers (like "WITHIN 300 SECONDS"). Figure 9 illustrates the structure of a STIX pattern.

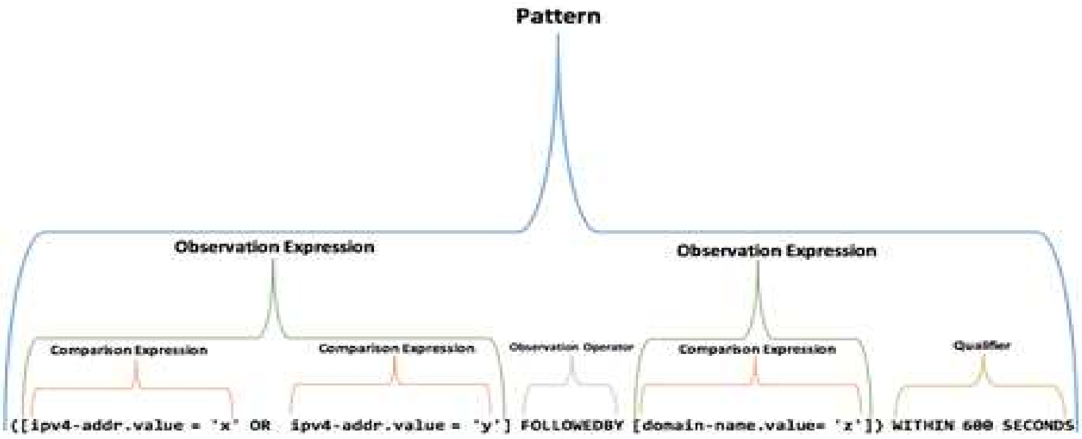


Figure 9. STIX Pattern Structure [57]

Example patterns are shown below, demonstrating the flexibility of STIX patterning, expressions can be nested and include constants in: text (default), binary (Base64 encoded), or hexadecimal:

Text:

```
[artifact:payload_bin = 'this is a test']
```

Binary:

```
[artifact:payload_bin = b'dGhpcyBpcyBhIHRlc3Q=']
```

Hex:

```
[artifact:payload_bin = h'74686973206973206120746573740']
```

Matching a File with a MD5 hash:

```
[file:hashes.MD5 = '79054025255fb1a26e4bc422aef54eb4']
```

Matching a File Object with a Windows file path:

```
[file:name = 'foo.dll' AND file:parent_directory_ref.path =
'C:\\Windows\\System32']
```

Matching on a URL:

```
[url:value = 'http://example.com/foo' OR url:value = 'http://example.com/bar']
```

Matching an Email Message with a particular From Email Address and Attachment File Name Using a Regular Expression:

```
[email-message:from_ref.value MATCHES '.+\\@example\\.com$' AND email-
message:body_multipart[*].body_raw_ref.name MATCHES '^Final Report.+\\.exe$']
```

Matching a File with a MD5 hash, followed by (temporally) a Registry Key Object that matches a value, within 5 minutes:

```
([file:hashes.MD5 = '79054025255fb1a26e4bc422aef54eb4'] FOLLOWEDBY [win-registry-
key:key = 'HKEY_LOCAL_MACHINE\\foo\\bar']) WITHIN 300 SECONDS
```

Matching on Network Traffic with a particular destination:

```
[network-traffic:dst_ref.type = 'ipv4-addr' AND network-traffic:dst_ref.value =
'203.0.113.33/32']
```

Matching, grouping with []:

```
[ipv4-addr:value = '198.51.100.1/32' OR ipv4-addr:value = '203.0.113.33/32'] AND
[network-traffic:dst_port = ‘22’]
```

5.1.3. Addressed issues and drawbacks

STIX 2.0 provides a rich format for expressing a broad range of CTI and complex relationships. STIX has been adopted by CTI vendors as a common standard, at this time the majority sampled have not yet upgraded from STIX v1.x presenting some incompatibility issues. The rich structure and subsequent size and complexity of STIX documents make the format cumbersome and inefficient when representing simple CTI, this is simpler to represent as a simple CSV text file or in a custom API JSON format.

5.2. TAXII (Trusted Automated Exchange of Intelligence Information)

Trusted Automated Exchange of Intelligence Information (TAXII) [55] is designed to be the preferred transport mechanism for retrieving and exchanging STIX data, it has been updated to v2.0, making compulsory the use of HTTPS. TAXII can be used as a RESTful API or by clients subscribing to published channels.

5.2.1. Use case

The use of TAXII for STIX is the preferred and not compulsory, other transports such as simple file download or custom API may be used. Neither does TAXII mandate that content should be in STIX format, the standard is open and may be used for any format of data.

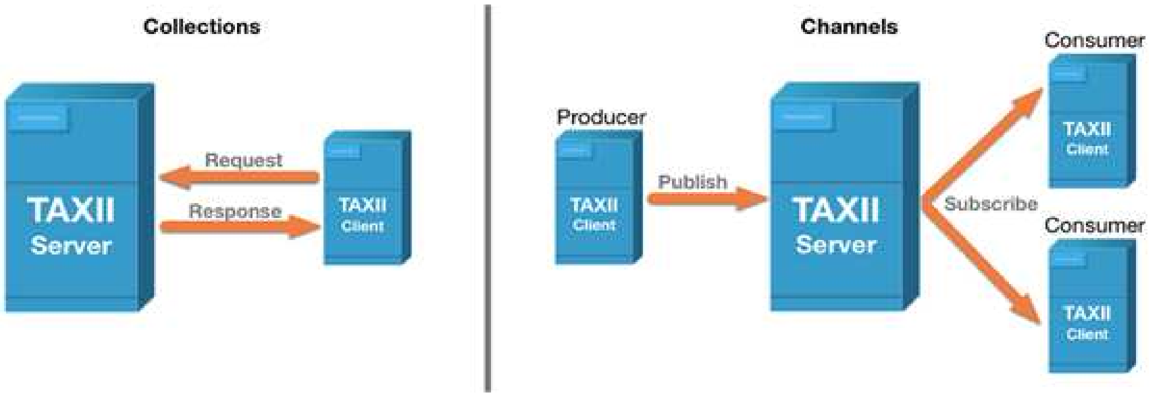


Figure 10. TAXII method of operation [55]

5.2.2. API and data model

- **Discovery**. The discovery of TAXII services can be performed by two methods represented in Figure 10 above:
  - Network level discovery uses DNS SRV records to advertise a TAXII server within a domain using the TAXII service name.
  - The second method is to publish the discovery endpoint URL by conventional means.
- **API Roots**. The TAXII server discovery service provides a list of API roots, each API root can contain one or more 'collections'.
- **Collection**. Collections contain metadata and one or more objects, where an object is typically a STIX JSON object. The TAXII specification allows the http header attributes to indicate and control

the range of content that should be downloaded in a request. URL parameters may be used to filter the content added after a given timestamp or for a given STIX: id, type, or object version.

- **Manifest**. The collection manifest lists all the objects available giving the: id, date_added, versions and media_types.
- **Channels**. The use and operation of subscription channels is still to be confirmed by the TAXII project.

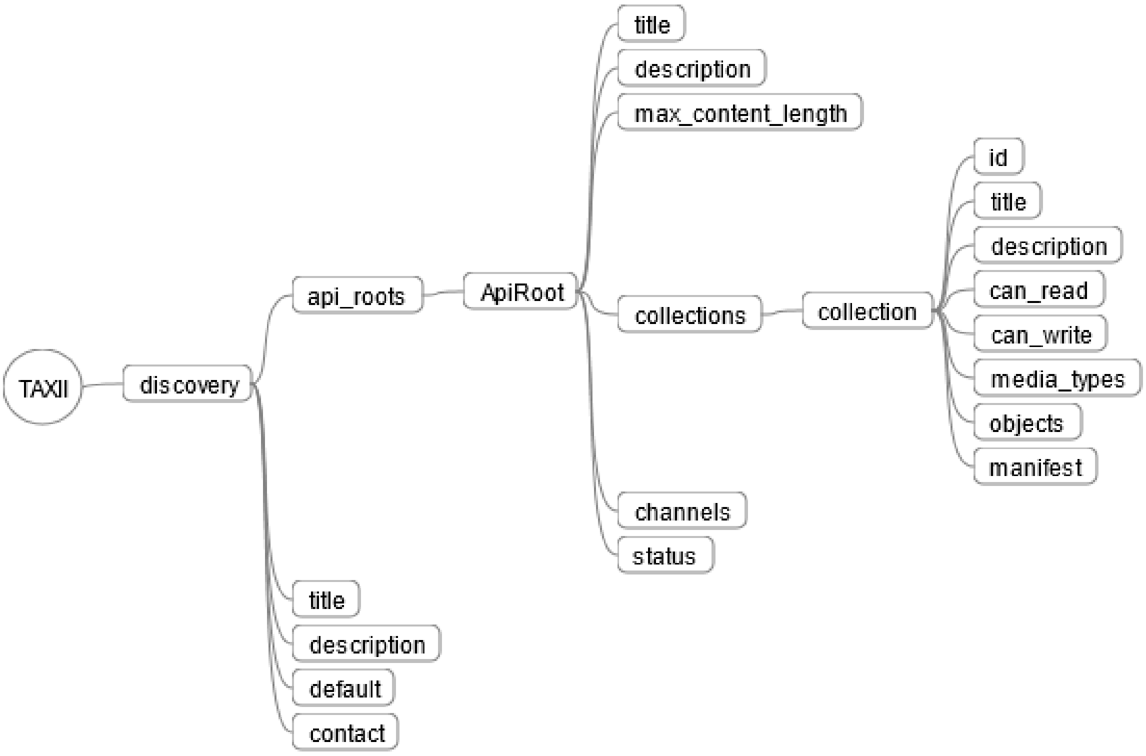


Figure 11. TAXII API

Example test script output for the Anomali TAXII service, below, shows three API Roots: 'TAXII feeds', 'Trusted Circles' and 'Search Filters'. Collections can be seen under the 'TAXII feeds'

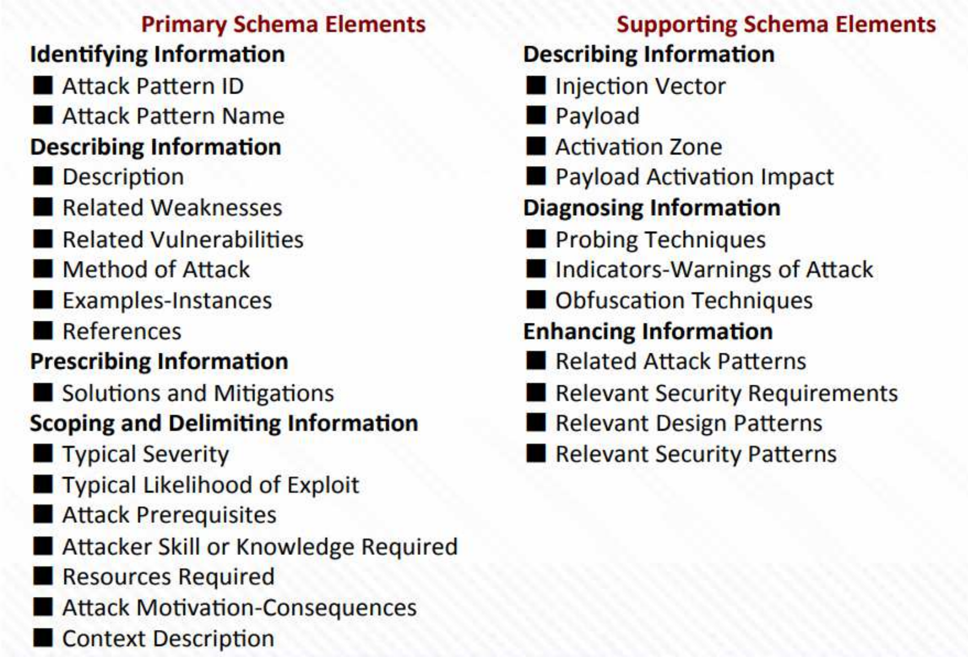
5.2.3. Addressed issues and drawbacks

TAXII is supported by several vendors, research found more of these were TAXII / STIX v1.x feeds than v2 feeds. The OASIS TAXII project supplies a Python library via GitHub for users to implement their own TAXII client, [58], or server, [59]. Alternatives include: IntelMQ, [10], and numerous platform-specific APIs, such as those provided by MSIP and CIF.

5.3. CAPEC (Common Attack Pattern Enumeration and Classification)

The Common Attack Pattern Enumeration Classification (CAPEC) is a dictionary that its goal is gathering information of intrusions and cyber-attacks and then classified the incidents and share the information by the open source platform to the experts for further analysis.
The main attributes of *CAPEC Dictionary* composed by *Mechanisms of Attacks* and *Domains of Attacks.* The *Mechanisms of Attacks* also consisting of information about the following topics: Collect and analyze

information, inject unexpected items, manipulate timing and state, abuse existing functionality, employ probabilistic techniques, manipulate data structures, manipulate system resources and etc. Also, in the *Domains of Attacks* we have domains originated from social engineering, supply chain, communications, software, physical security and hardware[5]. The above catalogue of CAPEC is providing schema elements that through use cases indicate a threat and through domino effects ends up indicating attack pattern and threat impact.



Primary Schema Elements

Identifying Information

- Attack Pattern ID
- Attack Pattern Name

Describing Information

- Description
- Related Weaknesses
- Related Vulnerabilities
- Method of Attack
- Examples-Instances
- References

Prescribing Information

- Solutions and Mitigations

Scoping and Delimiting Information

- Typical Severity
- Typical Likelihood of Exploit
- Attack Prerequisites
- Attacker Skill or Knowledge Required
- Resources Required
- Attack Motivation-Consequences
- Context Description

Supporting Schema Elements

Describing Information

- Injection Vector
- Payload
- Activation Zone
- Payload Activation Impact

Diagnosing Information

- Probing Techniques
- Indicators-Warnings of Attack
- Obfuscation Techniques

Enhancing Information

- Related Attack Patterns
- Relevant Security Requirements
- Relevant Design Patterns
- Relevant Security Patterns

Figure 12. General schema elements[6]

5.3.1. Data model

The background of CAPEC based on the technique of CWE. The goal of CWE is the improvement of software quality with respect to known security issues within source code. Also, it defines a common measurable in set of weaknesses and gives the opportunity for further description discussion, selection and utility of software security tools and services that can find these weaknesses. The unified consensus of CWE is to act or interact with taxonomies, products, perspectives, varying levels of abstraction included directory traversal, XSS variants, predominant in current research vocabulary, especially web application security.

To understand deeply how platform interacts with attacks, we should comprehend the general patterns where attack patterns are based on. Such general patterns are; attack/threat trees, fault trees, security patterns. Also, some of the attack patterns are; HTTP cookies, URL encoding, simple string injection and make the client invisible. When CAPEC framework and CWE method is combined the outcome performed in the below picture. The CWE method used is attack trees. Attack tree is a helpful diagram of relationship amongst asset actor-use case abuse case vuln-exploit countermeasure.

[5] https://capec.mitre.org/index.html

[6] https://interact.gsa.gov/sites/default/files/Mon AM2-SW Assurance Fall SSCA Forum-Sept 2015.pdf

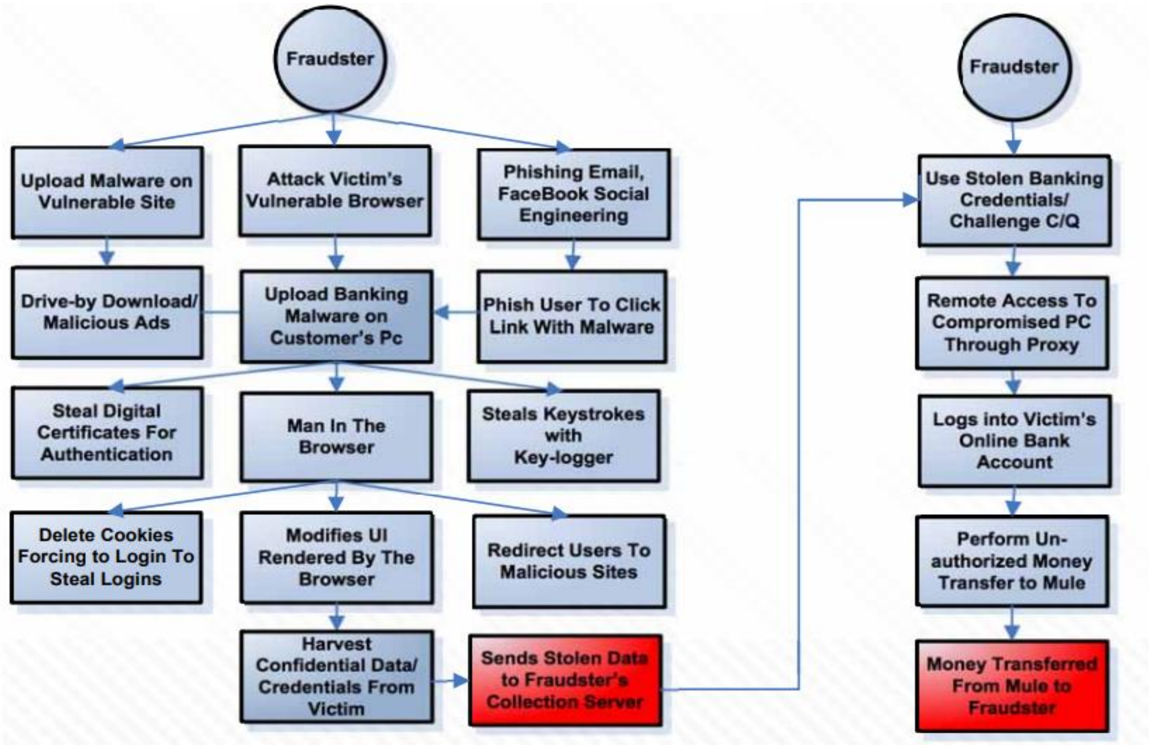

Fraudster
Upload Malware on Vulnerable Site
Attack Victim's Vulnerable Browser
Phishing Email, FaceBook Social Engineering
Drive-by Download/ Malicious Ads
Upload Banking Malware on Customer's Pc
Phish User To Click Link With Malware
Steal Digital Certificates For Authentication
Man In The Browser
Steals Keystrokes with Key-logger
Delete Cookies Forcing to Login To Steal Logins
Modifies UI Rendered By The Browser
Redirect Users To Malicious Sites
Harvest Confidential Data/ Credentials From Victim
Sends Stolen Data to Fraudster's Collection Server
Fraudster
Use Stolen Banking Credentials/ Challenge C/Q
Remote Access To Compromised PC Through Proxy
Logs into Victim's Online Bank Account
Perform Un-authorized Money Transfer to Mule
Money Transferred From Mule to Fraudster

Figure 13. Analysis of attacks using attack trees

5.4. MAEC (Malware Attribute Enumeration and Characterization)

The *Malware Attribute Enumeration and Characterization* (MAEC) language is designed for characterizing malware using attributes such as behaviors, artifacts, and relationships between malware samples. [38]. The current version 5 release has been updated in line with STIX v2.0 to maintain compatibility. The principal changes are a simplified structure, JSON serialization, RFC7159, eases the encoding and decoding of data for programmers as it is a direct representation of in-program JavaScript objects, Python dictionaries, Perl hashes, etc. MEAC leverages STIX observable objects and types to provide commonality between the two formats.

5.4.1. Use case

MAEC provides a common language to characterize malware, this can reduce ambiguity in malware descriptions and duplication. Sandbox resources such as the VxStream Sandbox, [29]' which amongst other formats, provides the results from malware analysis in MAEC format (limited access after free registration and vetting, paid-for full access available).

5.4.2. Data model

MAEC uses a graph-based data model with five top-level objects: Behaviors, Collections, Malware Actions, Malware Families and Malware Instances. These along with their relationships, including STIX observable objects are depicted below [41].

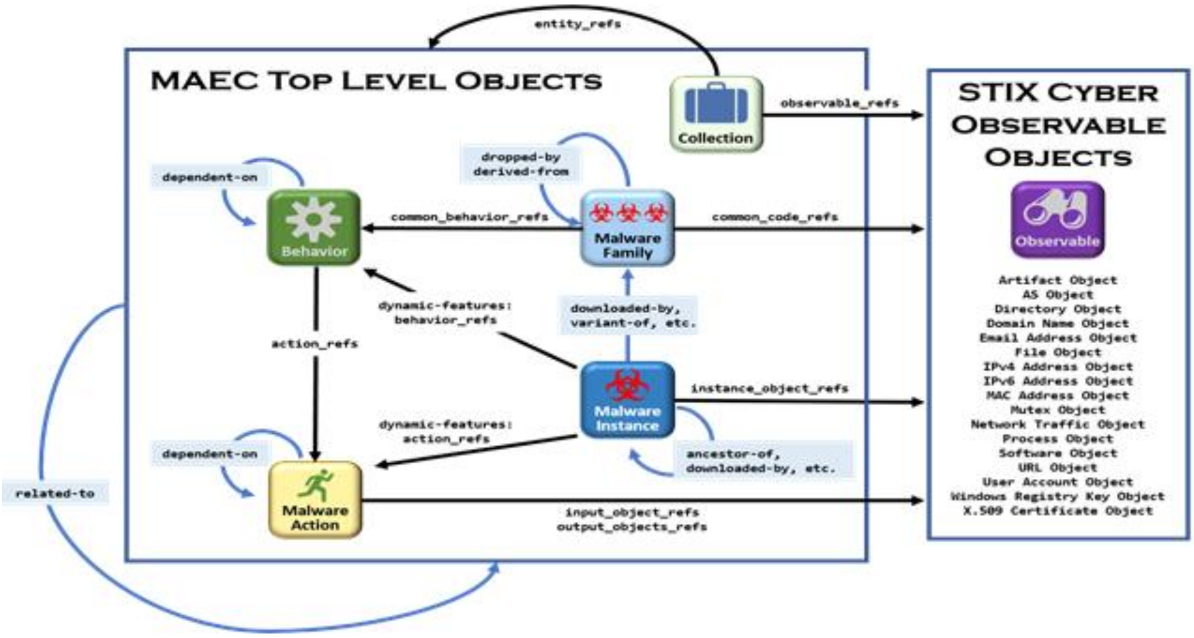


Figure 14. MAEC top level objects

The MAEC package data model is shown below, the root element of the document is always “package” which must contain the “type” (always “package”), id in the format “package-UUID”, “schema_version” is “5.0”. The “maec_objects” array must contain one or more MAEC objects describing the malware. The STIX “observable_objects” are optional. The MAEC data model is documented in the “MAEC Vocabularies Specification” [42].

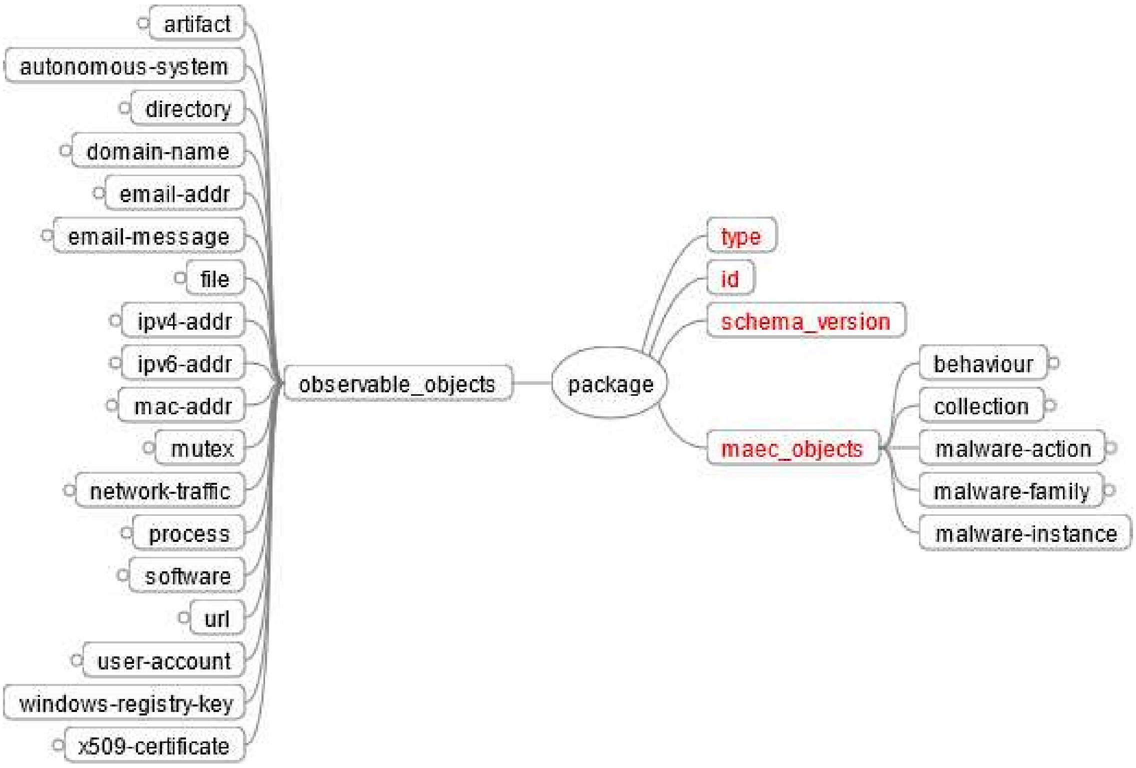


Figure 15. MAEC Data model

As an example, in case a MAEC package includes a “malware-instance” type of object, this can reference the observed details in the STIX “observable_objects” structure, which are specified in arrays like “instance_object_refs” and “tool_refs”. This will provide a high level of detail about an executable file carrying a malicious payload and information concerning the tools used in the analysis.

5.4.3. Addressed issues and drawbacks

The MAEC format is supported by some sandbox providers, (JoeSandbox Pro & VxStream), can both provide MAEC 4.1 format downloads. All sandbox providers examined, e.g., VirusTotal, provide their own JSON format via API or download access. The MAEC project have written a Python tool to download VirusTotal reports and convert them to MAEC 5.0; available on the project GitHub repository.

5.4.4. MAEC and STIX

MAEC and STIX were designed to serve different roles. However, MAEC content can also be embedded in STIX permitting the two languages to complement each other. In such a case, they permit the capture of detailed malware information along with related cyber-threat information. This allows for finer-grained relationships between malware and the larger cyber-threat context to be established and expressed; this is depicted in the following table, against a number of features related to the project, where the mark ‘?’ (resp. ‘X’) indicates that the feature is partially (resp. fully) provided.

Table 26. Joint benefits of STIX and MAEC

Feature	STIX	MAEC
IP address	X	X
Email address	X	X
Hostname	X	X
URL/URI	X	X
Domain	X	X
Attacker/target	X	X
Vulnerability	?	X
Malware/threat type	?	X
Ransomware		X
File details	X	X
System IoCs	X	X
DDoS		X
Spam		X
Phishing		X
Software	X	X
Time stamps	X	X
CTI source	X	X
Rich CTI data	X	X
Patterns	X	X
Identity	X	
Course of action	?	
Versioning	X	
Author	X	
Confidence/count	X	

Compromised host		?
Botnet		X
Markings	X	
Artifacts	X	X

5.5. IDMEF (Intrusion Detection Message Exchange Format)

The *Intrusion Detection Message Exchange Format* (IDMEF) defines data formats and exchange procedures for sharing information relating to intrusion detection and response systems and to the management systems that may need to interact with them. This format enables interoperability among commercial, open source, and research systems, so users can mix-and-match the deployment of these systems according to their strong and weak points to obtain an optimal implementation.

5.5.1. Applicability of IDMEF

The most obvious place to implement the IDMEF is in the data channel between an intrusion detection analyzer and the manager to which it sends alarms. Still, there are other places where the IDMEF can be useful, such as:

- In a **single database system**, to store results from different intrusion detection products, in order to analyze data in their completeness.
- In an **event correlation system**, that accepts alerts from a variety of intrusion detection products. This will result in more sophisticated cross-correlation and cross-confirmation calculations.
- In a **graphical user interface**, to display alerts from a different intrusion detection products, to allow the user to monitor all of the products from a single screen, and require him to learn only one interface, instead of several.
- In **different organizations** in general (users, vendors, response teams, law enforcement), may use IDMEF to not exchange data, but also to comment and evaluate them.

5.5.2. Rationale for implementing IDMEF in XML

XML's flexibility makes it a good choice for implementing the IDMEF. In more detail, XML allows a custom language to be developed specifically for the purpose of describing intrusion detection alerts. It also defines a standard way to extend the developed language, either for later revisions of the IDMEF (standard extensions) or for vendor-specific use (non-standard extensions). Additionally, software tools for processing XML documents are widely available, in both commercial and open source forms. Numerous tools and APIs for parsing and validating XML are available in a variety of languages, including Java, C, C++, Tcl, Perl, Python, and GNU Emacs Lisp. Widespread access to tools makes the adoption of IDMEF by product developers easier and faster. One of IDMEF's requirements, is that message formats support full internationalization and localization. XML meets that requirement, as the XML standard requires support for both the UTF-8 and UTF-16 encodings of ISO/IEC 10646 and Unicode, making all XML applications (and therefore all IDMEF-compliant applications) compatible with these common character encodings. Another IDMEF requirement, is that message formats must support filtering and aggregation. XML's integration with the style language XSL, allows messages to be combined, discarded, and rearranged. Thus, XML meets this requirement as well. Ongoing XML development projects, in the W3C and elsewhere, provide object-oriented extensions, database support, and other useful features. With an XML implementation, the IDMEF immediately also utilizes such features. Finally, XML is free, without license, fees, and royalties.

5.5.3. Data model

The IDMEF data model is an object-oriented representation of the alert data sent to intrusion detection managers by intrusion detection analyzers. Depending on the capabilities of the analyzer that creates an alert, the alert may be simple or complex. The data model was designed to provide a standard representation of such alerts in an unambiguous fashion, and to permit the relationship between simple and complex alerts to be easily described.

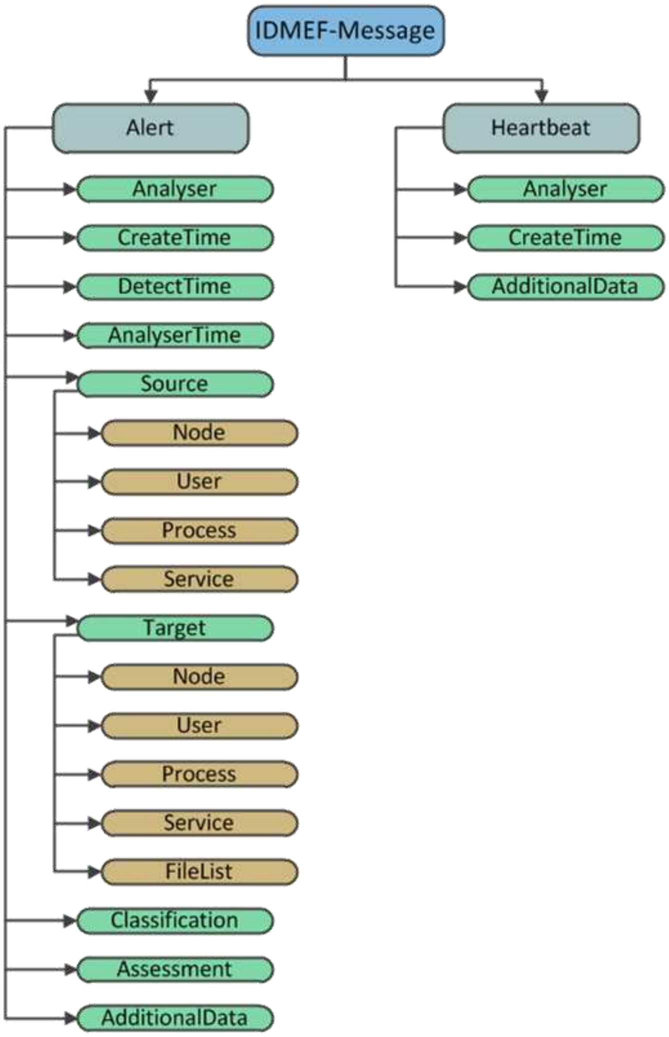


Figure 16. The IDMEF data model

The design of the data model is content-driven; new objects are introduced to accommodate additional content rather than semantic differences between alerts. This is very important, as the task of classifying and naming computer vulnerabilities is both extremely difficult and very subjective. The data model is unambiguous; this means that while the analyzers are allowed to be precise, they are not allowed to produce contradictory information in two alerts describing the same event.

5.5.4. Problems addressed by the data model

The IDMEF data model at hand addresses several problems associated with representing intrusion detection alert data. Below we briefly describe the problems and outline the benefits of IDMEF with respect to these problems.

- **Alert information is heterogeneous.** Some alerts are defined with very little information, such as origin, destination, name, and time of the event. Other alerts provide additional information, such as ports or services, processes, user information, etc. The data model is flexible enough to accommodate different needs. An object-oriented model is naturally extensible via aggregation and subclassing that provide extensibility while preserving the consistency of the model.
- **Intrusion detection environments are different.** Some analyzers detect attacks by analyzing network traffic. Others use operating system logs or application audit trail information. Alerts for the same attack, sent by analyzers with different information sources, will not contain the same information. The data model defines support classes that accommodate the differences in data sources among analyzers.
- **Analyzer capabilities are different.** Depending on the environment, one may install a lightweight analyzer that provides little information in its alerts, or a more complex analyzer that will have a greater impact on the running system but provide more detailed alert information. The data model defines extensions to the basic Document Type Definition (DTD) that allow carrying both simple and complex alerts. Extensions are accomplished through subclassing or association of new classes.
- **Operating environments are different.** Depending on the kind of network or operating system used, attacks will be observed and reported with different characteristics. The data model accommodates these differences with the Node and Service support classes. If additional information must be reported, subclasses can be defined that extend the data model with additional attributes.
- **Commercial vendor objectives are different.** For various reasons, vendors may wish to deliver more or less information about certain types of attacks. The object-oriented approach allows this flexibility while the subclassing rules preserve the integrity of the model.

5.6. IODEF (Incident Object Description Exchange Format)

In order to address the issue of the immediate exchange of information about computer security incidents between CSIRTs, the *Incident Object Description Exchange Format* (IODEF) defines a data representation that provides a proper framework. The main idea behind the implementation of IODEF by CSIRTs is the improvement of their operational capabilities, since it provides possibilities such as:

- The increase of the automation in incident data processing.
- The decrease of the normalization effort of similar data from different sources.
- The development of interoperable tools for incident handling and subsequent analysis, through the adoption of that common format.

5.6.1. The IODEF data model

One of the main characteristics behind the development of the IODEF data model, is its compatibility with the preceding one; the Intrusion Detection Message Exchange Format (IDMEF) data model, which was developed for intrusion detection systems. For this reason, the IODEF is mainly based on the IDMEF, and it provides backward compatibility with it.

The IODEF rationale is to define a structured format, consisted of all valuable information about computer security incidents, with a view to creating a common communication code between all involved CSIRTs. Such information may be: the identification of the incident, other incidents caused by the one identified, the time of its detection, its duration and when it was reported, the techniques followed by the intruder during it, etc.

Deriving from its rationale, the IODEF data model describes an IODEF-Document, which is the top level class in the IODEF data model and each IODEF document is an instance of this class. The IODEF-Document class contains one or more Incident classes, which represent every incident. The Incident class specifies a standard representation for commonly exchanged incident data. Thus, in the IODEF model, the classes that constitute Incident class, contain all valuable information about the incident, and they are:

- **IncidentID:** One. An incident identification number assigned to this incident by the CSIRT who generates the IODEF document.
- **AlternativeID:** Zero or One. The incident ID numbers used by other CSIRTs to refer to the incident described in the document.
- **RelatedActivity:** Zero or One. The ID numbers of the incidents which are linked to the one specified in this document.
- **DetectTime:** Zero or One. The time when the incident was first detected.
- **StartTime:** Zero or One. The time when the incident started.
- **EndTime:** Zero or One. The time when the incident ended.
- **ReportTime:** One. The time when the incident was reported.
- **Description:** Zero or More. A non-formatted textual description of the event.
- **Assessment:** One or More. A characterization of the incident's impact.
- **Method:** Zero or More. The techniques followed by the intruder during the incident.
- **Contact:** One or More. Contact information for the parties involved in the incident.
- **EventData:** Zero or More. Description of the events comprising the incident.
- **History:** Zero or One. A log of notable events or actions that happened during the course of handling the incident.
- **AdditionalData:** Zero or More. A mechanism that helps extending the data model.

Additionally, the IODEF implementation is described with an Extensible Markup Language (XML) Schema and it provides numerous advantages. It is ideal for specifying a data encoding framework that supports various character encodings. Furthermore, the plethora of related technologies (e.g., XSL, Xpath, XML-Signature) simplifies the IODEF documents manipulation. However, it is ineffective when exchanging large volumes of data or embedding binary data, since XML is essentially a text representation.

The IODEF data model addresses several problems in representing incident data:

- Incident data is dissimilar. Following an object-oriented approach, the model provides extensibility by aggregating and creating additional sub-classes, preserving its consistency. So, if the data model required modification, it is extended with new classes. It doesn't cause any incompatibility issues for implementations which do not recognize these extensions, since the basic subset of the data model is still understood.
- Incidents have a life-cycle, which leads to having different information or levels of detail be present according to their stage in the cycle. The data model is flexible to accommodate such different needs.

- Communication and coordination are the main activities of the involved CSIRTs. The data model provides the definition of support classes that accommodate information about the incident data reporting sources, in order to keeping track of the reporter's identity as well as to specifying a confidence level to the submitted information.
- Incident data may encompass confidential information which shouldn't be exposed to unauthorized parties during collaboration. The data model allows the indication of restrictions on the usage of the data through a granular tagging in the individual classes. However, the incident handling system, implementing the data model, specifies these labels.

Finally, in the design of the IODEF data model, a number of considerations were made. The data model represents a transport format and hence, its representation is not optimal for on-disk storage or in-memory processing. Next, since there is no precise definition widely agreed for an incident, the data model does not attempt to prescribe one through its implementation. Rather, the IODEF is flexible enough to cover most operators' needs. Moreover, as it would require an extremely complex data model to describe an incident for all definitions, and the IODEF is only designed to be a framework for conveyance of commonly exchanged incident information, it is ensured that there are sufficient mechanisms for extending the model to support specific organization needs for information. Lastly, since the domain of security analysis is not fully standardized and relies on unstructured textual descriptions, the IODEF attempts to be a balancing factor between the support of this free-form content and the allowance of automated processing of incident information.

5.6.2. Example of an incident reporting following the IODEF data model

In the following example, a CSIRT reports an incident on 2006-06-08, 05:44:53, about a large botnet named GT Bot, exploiting the CA-2003-22 vulnerability to be installed in the system. The coordinator of this incident is Joe Smith, and his contact information is: jsmith@csirt.example.com (email). There are two bots running in 192.0.2.1 and 192.0.2.3, and they are sending DoS traffic at 10 KB/sec and 250 KB/sec accordingly. These bots communicate with the IRC server on #give-me-cmd channel (irc.example.com running on 192.0.2.20). In order to resolve this high severity issue, the CSIRT expects these actions to be followed: Confirm the sources, take the machines offline and remediate.

```xml
<IODEF-Document version="1.00" lang="en"
  xmlns="urn:ietf:params:xml:ns:iodef-1.0"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemalocation="urn:ietf:params:xml:schena:iodef-1.0">
  <Incident purpose="mitigation">
       <IncidentID name="csirt.example.com">908711</IncidentID>
       <ReportTime>2006-06-08T05:44:53-05:00</ReportTime>
       <Description>Large botnet</Description>
       <Assessment>
             <Impact type="dos" severity="high" completion="succeeded" />
       </Assessment>
       <Method>
             <!-- References a given piece of malware, "GT Bot" -->
             <Reference>
                    <ReferenceName>GT Bot</ReferenceName>
```

```
        </Reference>
        <!-- References the vulnerability used to compromise the machines -->
        <Reference>
            <ReferenceName>CA-2003-22</ReferenceName>
            <URL>http://www.cert.org/advisories/CA-2003-22.html</URL>
            <Description>Root compromise via this IE vulnerability to install
the GT
                Bot</Description>
        </Reference>
    </Method>
    <!-- A member of the CSIRT that is coordinating this incident -->
    <Contact type="person" role="irt">
        <ContactName>Joe Smith</ContactName>
        <Email>jsmith@csirt.example.com</Email>
    </Contact>
    <EventData>
        <Description>These hosts are compromised and acting as bots
            communicating with irc.example.com.</Description>
        <Flow>
            <!-- bot running on 192.8.2.1 and sending DoS traffic at 10KB/sec -
->
            <System category="source">
                <Node>
                    <Address category="ipv4-addr">192.0.2.1</Address>
                </Node>
                <Counter type="byte" duration="second">10000</Counter>
                <Description>bot</Description>
            </System>
            <!-- a second bot on 192.0.2.3 -->
            <System category="source">
                <Node>
                    <Address category="ipv4-addr">192.0.2.3</Address>
                </Node>
                <Counter type="byte" duration="second">250000</Counter>
                <Description>bot</Description>
            </System>
            <!-- Command-and-control IRC server for these bots -->
            <System category="intermediate">
                <Node>
                    <NodeName>irc.example.com</NodeName>
                    <Address category="ipv4-addr">192.0.2.20</Address>
                    <DateTime>2006-05-08T01:01:03-05:00</DateTime>
                </Node>
                <Description>IRC server on #give-me-cmd
channel</Description>
            </System>
        </Flow>
        <!-- Request to take these machines offline -->
```

```
             <Expectation action="investigate">
                   <Description>Confirm the source and take machines offline and
                         remediate</Description>
             </Expectation>
       </EventData>
  </Incident>
</IODEF-Document>
```

5.7. CVRF (Common Vulnerability Reporting Framework)

The *Common Vulnerability Reporting Framework* (CVRF) (https://www.icasi.org/cvrf/) is an XML-based format derived from IODEF that addresses issues of exchanging vulnerability information and other security related documentation [12]. The key insight behind CVRF is its ability to automate and homogenize the creation and consumption of vulnerability documentation, so it can provide the following advantages:

- It can provide a consistent way to disseminate security information, by simplifying the interpretation of the advisories.
- It can provide a machine-readable format for the interpretation of security advisories, by allowing automation.

CVRF allows sharing critical security information among various stakeholders, in just a single format, to speed up information possessing. CVRF's rationale is that when a document producer (CERTs, security firms, researchers, etc.) creates a vulnerability report, it will use an automated, common, and expected format that document consumers (vendors, end-users, etc.) can parse and understand.

5.7.1. The CVRF Data model

Different versions of CVRF was deployed in the past few years: the ICASI CVRF v1.0 in 2011, a year later CVRF v1.1 [12], and the latest CVRF v1.2 [14] in 2017, which was released from the OASIS *Common Security Advisory Framework* (CSAF) technical committee, that is not backward compatible with the older versions. The CVRF data model is a graphical tree rendering the relations among the elements under the single root, which is provided in the following diagram.

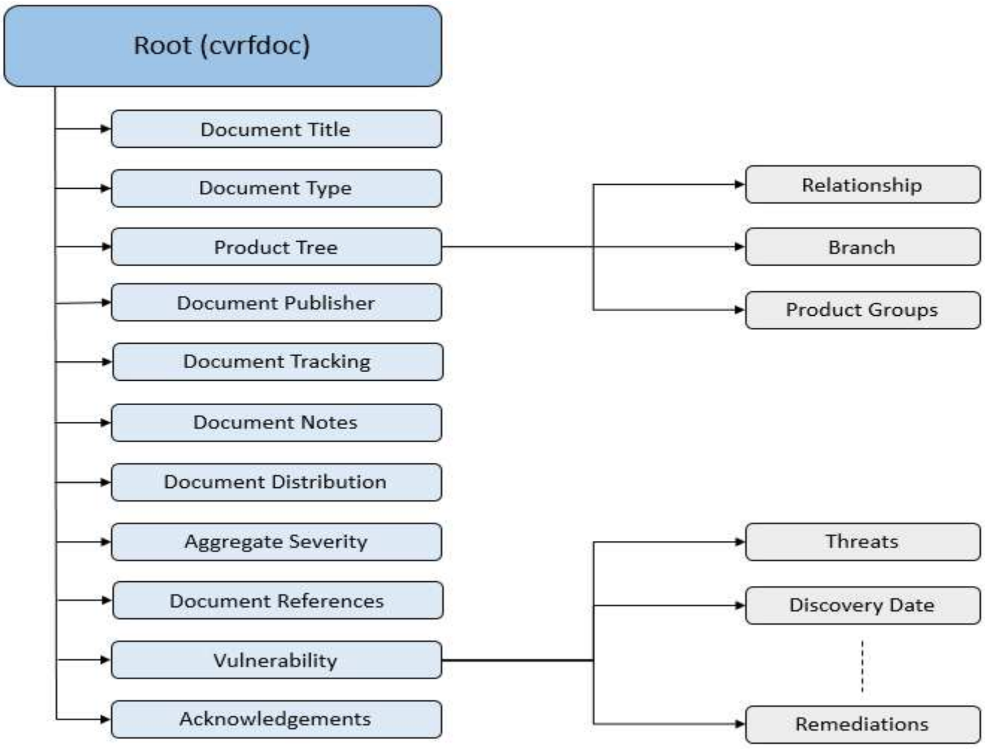


Figure 17. The CVRF data model

Amongst the various element types, those related to the CVRF document (i.e., the “document title”, the “document type”, the “document publisher”, and the “document tracking”) are mandatory, whereas the other are optional. All the elements can appear at most once in the document, with the exception of the “vulnerability” element that provides detailed information about threats (including possible remediation actions).

5.7.2. Adoption of CVRF by organizations

Due to the need for automating security and allow for more sophisticated and dynamic ways to patch and configure the available systems CVRF has been adopted by several organizations and large industry players, including Cisco, IBM, Intel, Juniper, Microsoft, Nokia, Oracle, and Red Hat.

5.7.3. Addressed issues and drawbacks

- **Design**. One of the major decisions made for the design of the CVRF was to adhere to a data driven model focusing on how to package up and structure the data, without the concern of how this information will be presented to the end user. Additionally, CVRF is just an XML language, which means a structured way for users to create, store and trade information. The way that this information is going to be parsed, used and transported is in short words up to the users.
- **Data character**. CVRF has considerably enough unrestricted (string-typed) fields that are human readable. That means, that any document producer could input arbitrary data without any constraint of the content or the length. Although, there are cases where a document producer is forced to insert data that contains markup into a CVRF document. In such cases, an XML parser is advised to transform the document.

- **Product tree**. Was firstly included in CVRF 1.1 [13], and totally redesigned the way that CVRF enumerates products in a vulnerability. The way that the products were specified and created several issues to the CVRF users. In case where a product had enough vulnerabilities, it was encapsulated into a Vulnerability container several times, repeating by this way the same XML code. Using the Product Tree was liberating because a reference of the same product could be inserted in each Vulnerability container rather than enforced to list each product again and again.

Summarizing, the benefits using CVRF is that the end users will be able to process, find and act upon security information quickly and easily, with a higher level of confidence that the information is accurate and comprehensive. A long-term goal is that CVRF will make it easier for researchers to submit vulnerability issues to companies, which is a process that has been fraught with complexity until now. The capability of being able to submit data using an XML-based system that allow security information to circulate more rapidly than has ever happened in the past.

5.8. OVAL (Open Vulnerability and Assessment Language)

The *Open Vulnerability and Assessment Language* (OVAL) [62] is a worldwide security-based community which promotes open security content publicly available through a plethora of security tools and services. OVAL (https://oval.mitre.org/language/version5.11/) contains a language which determines specific steps of the assessment process. Firstly, represents configuration information of the system, then analyzes it for the presence of the specified machine state (configuration, vulnerability, patch state, etc.); and finally reports the results of this assessment. The repositories are collections of publicly available and open content that utilize the language.

OVAL is comprised of three different XML-based schemas to serve as the vocabulary and the framework of the language. These different categories: system results, definitions and characteristics corresponds to the assessments process and each one varies of component schemas.

5.8.1. Use cases

Interoperability is enabled by the OVAL language between security products by using an XML format to exchange information, meaning that tools can leverage other tools which belongs to different vertical markets and accomplish different tasks (e.g., a vulnerability assessment product could leverage a vulnerability research service in order to check the latest vulnerabilities quickly and automatically). Therefore, we present the potential use cases of the OVAL language [62]:

- **Security advisory distribution**: The security advisories contain all the needed information for researchers and vendors to discover product vulnerabilities on a system. All this information is published in a machine-readable format.
- **Vulnerability assessment** is a process to identify vulnerabilities in a system and according to their severity can set them in a priority list. Vendors can quickly and easily consume data by having the information stored in a standard format and from the tool's customers view. This format makes it more easy to determine which tool has the best features for their needs.
- **Patch management** is a process to identify security issues on a system as well as software updates that affects it. The assurance that a system is patched by the proper way is a major concern to the patch management vendors because they are able to consume data from multiple sources and the interoperability between tools is no longer affected.

- **Configuration management** is a process where a machine's configuration state is examined either by a normal profile built in the system or a profile created by a programmer and then reporting the results.
- **Auditing and centralized audit validation** which is responsible for providing the needed reports at any time in the past for any machine in the system. Firstly, the capturing of the machine configuration information and then the storing of that information is needed in a standard, data centric format ensuring that it is not only stored in a specific tool which may not be available if someone need to review the data.
- **Security Information Management Systems (SIMS)** may use various security products and agents to create a completed view of the security of an organization. A product can be powerful and flexible if the data needed by the SIM is as few as possible.
- **System Inventory is the operation of assembling** in a list all the applications installed in a system. Many organizations have different versions of applications running on a wide variety of operating systems and need to track software for licensing and vulnerability management. So the need to check which applications and which versions of them are installed in a system is formed.

5.8.2. The OVAL structure

The OVAL Definition Schema is addressed to define a framework needed to write:

- The Vulnerability Definitions by determining the circumstances needed to exist if a certain vulnerability is going to be presented.
- The Patch Definitions to determine if a patch is suitable for our system.
- The Inventory Definitions to define all the conditions which determine whether a software is installed on the system.
- The Compliance definitions which define compliance with a statement or a policy.

The OVAL System Characteristic Schema, that represents the system's configuration information which includes installed software applications, OS parameters and other security-based values [62]. This schema provides a database of characteristics which are compared to the OVAL Definitions analyzing this way a system for patches, vulnerabilities and configuration issues.

The OVAL Results Schema determines an XML format [62] which is used to store the evaluation results of the system [62]. This data represents the system's configuration state compared to a set of OVAL Definitions. This schema allows the consummation of the data, the interpretation and usually takes the appropriate actions to mitigate the system's vulnerabilities and configuration issues (e.g., alter the configuration settings, take precautions to restrict the access to affected systems, install patches). The OVAL Results Schema's goal is to exchange a vulnerability and configuration format that is incorporated into a variety of OVAL compatible products and tools.[7]

[7] https://oval.mitre.org/compatible/compatible.html.

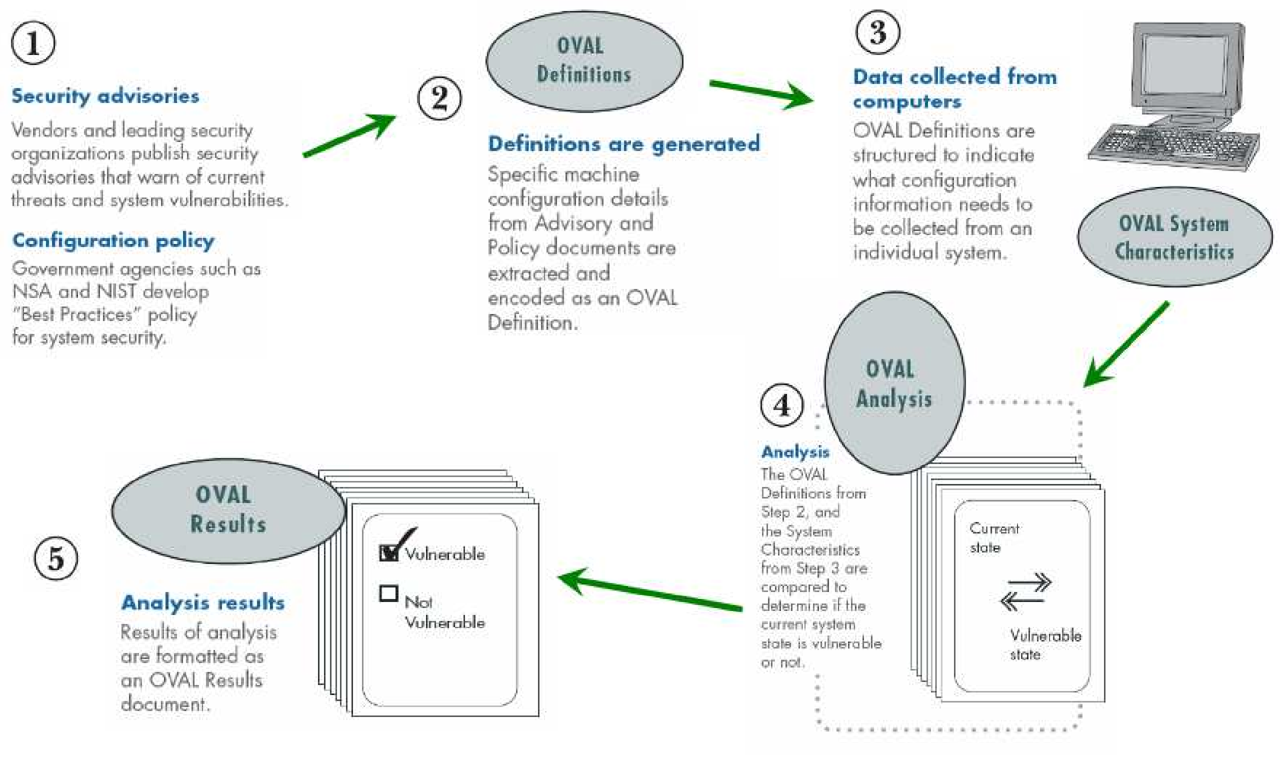
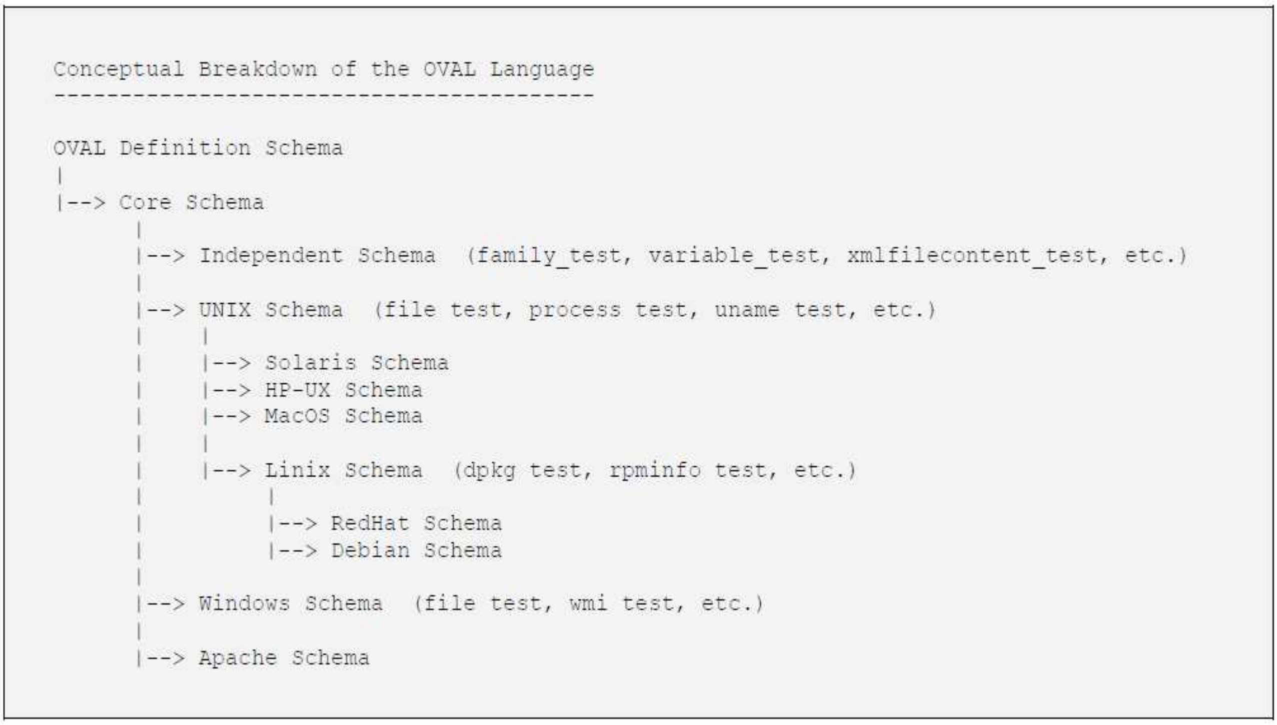

1
Security advisories
Vendors and leading security organizations publish security advisories that warn of current threats and system vulnerabilities.
Configuration policy
Government agencies such as NSA and NIST develop "Best Practices" policy for system security.
2
OVAL Definitions
Definitions are generated
Specific machine configuration details from Advisory and Policy documents are extracted and encoded as an OVAL Definition.
3
Data collected from computers
OVAL Definitions are structured to indicate what configuration information needs to be collected from an individual system.
OVAL System Characteristics
4
OVAL Analysis
Analysis
The OVAL Definitions from Step 2, and the System Characteristics from Step 3 are compared to determine if the current system state is vulnerable or not.
Current state
Vulnerable state
5
OVAL Results
Analysis results
Results of analysis are formatted as an OVAL Results document.
☑ Vulnerable
□ Not Vulnerable

Figure 18. How OVAL works[8]

These three schemas are not individual schemas but each one is comprised of a “collection of components” and a “core schema”. The function of the Definition’s core schema which is independent of the tests, provides a place to express the metadata (e.g., affected platforms, descriptions) while the component schema determines the tests used by the OVAL Language to identify security issues and vulnerabilities which affect an operating system. In the System Characteristic Schema, the function is similar to the Definition Schema, but the component schema determines the content and the format of the configuration parameters that are collected and corresponded to the tests determined within the OVAL Definition schema. Once again, the Results schema is a composed of a component schema and a core schema too. The difference between the Results schema and the Definition is that the OVAL Results schema in each test contains the information used to determine the existence of a vulnerability or a misconfiguration with the analyzed data in order to arrive at this result.

The OVAL Component Schemas [62], are comprised of a set of OVAL tests that are associated with the software they describe. The Tests are indistinguishable across the different schemas and they are grouped as a higher-level component schema, which means that all have the same structure for a file Test. Due to this similarity, a file Test is ordered in a conceptually higher-level UNIX schema.

[8] http://oval.mitre.org/documents/docs-06/an_introduction_to_the_oval_language.pdf



```
Conceptual Breakdown of the OVAL Language
-----------------------------------------

OVAL Definition Schema
|
|--> Core Schema
      |
      |--> Independent Schema  (family_test, variable_test, xmlfilecontent_test, etc.)
      |
      |--> UNIX Schema  (file test, process test, uname test, etc.)
      |     |
      |     |--> Solaris Schema
      |     |--> HP-UX Schema
      |     |--> MacOS Schema
      |     |
      |     |--> Linix Schema  (dpkg test, rpminfo test, etc.)
      |          |
      |          |--> RedHat Schema
      |          |--> Debian Schema
      |
      |--> Windows Schema  (file test, wmi test, etc.)
      |
      |--> Apache Schema
```

Figure 19. An OVAL hierarchical structure[9]

The OVAL hierarchical structure has no directed links between schemas. For example, as shown in Figure 19, the MacOS Schema does not inherit the aspects of the from the package test of the Unix Schema, but all tests inherit a collection of attributes and a note element from the core schema. Using this structure, the Definitions could grab tests from different schemas.

5.8.3. Addressed issues and drawbacks

As mentioned earlier the separation of tests where different types of software can be described (e.g., tests that are associated with Solaris and RED Hat), while simultaneously are gathering related tests together (e.g., tests associated with UNIX aspects). In case where a software vendor needs to create an addition to support the OVAL Language, he could gather a set of new tests into a new component schema instead of modifying any existing one.

Dissecting of the OVAL language into three schemas allows tools to target a number of applications and OSs directly as consequence to diminish their execution time and the process overhead [62]. In earlier versions, tools were focused on managing a single schema that may sometimes include tests irrespective to the tool, and by this way increase the processing overhead.

Addressing to the negative issue of this approach is that such schemas could be difficult or unwieldy for users to navigate. To address that though, we must ensure that the rules that govern the schema's structure must be well formed, followed consistently and well documented and thus provide the user the ability to find any test in the schema. Summarizing, we list some of the benefits gained by using OVAL:

- A unified approach to see if there is a vulnerability, patch or configuration issue in a system.

[9] https://oval.mitre.org/language/about/structure.html

- A single XML-based document encoding the details (with sufficient accuracy) of that specific issue.
- Precise schemas that emphasize the essential configuration information.

5.8.4. Adoption by product list

OVAL is implemented by a number of products and organizations in order to support functionalities, such as vulnerability assessment, configuration management, auditing, network connection health checking, and patch management, and risk management; a non-exhaustive list of examples includes Altex-Soft, CISCO, Greenbone Networks, jOVAL.org, McAfee, IT Promotion Agency, NopSec, OpenVAS, GCP Global, Positive Technologies CJSC, Red Hat, Beyond Trust, and SAINT Corporation, amongst others.

5.9. OPENC2 (Open Command and Control)

The *Open Command and Control (OPENC2)*[10] is a framework that its goal is the coordination of defense in cyber-relevant time. This will be achieved with the following tactics: (a) decoupling functional blocks and standardizing interfaces; (b) identifying and filling the gaps since they pertain to CTI indicator sharing and response; and (c) participating in a diverse and collaborative environment.

5.9.1. Decoupled security stack

The decoupled security stack is the separation of the whole system into 4 regulating blocks (sensing, sense making, decision making, acting). Substantially, the Decoupled Security Stack make the process of OPENC2 more manageable and easier to be understand, due to the whole system is breaking in 4 pieces and each piece integrated separately. Nevertheless, in the final stage the information is coming together. OPENC2 characterized by a) Unambiguous Machine-to-Machine Communication, b) Simplicity by Low overhead on sensor and actuator, c) Focuses on 'Acting' portion of cyber defense. More specific, OPENC2 assumes that the sensing, sense making and decision making regulating blocks have already been done before the acting part starts d) OPENC2 will leverage pre-existing protocols and efforts.

OPENC2 is a lightweight source, since has efficient machine-to-machine communications, is extensible, since its format extensions enables additional precision and flexibility. Defined as agnostic and transport because of its authentication, integrity controls as well as enable flexibility with respect to implementation. Last but not least, a significant characteristic is the abstract which focuses on 'What' to do and do not describe the specific object (device).

[10] https://www.oasis-open.org/events/sites/oasis-open.org.events/files/Wednesday-Session-5.pdf

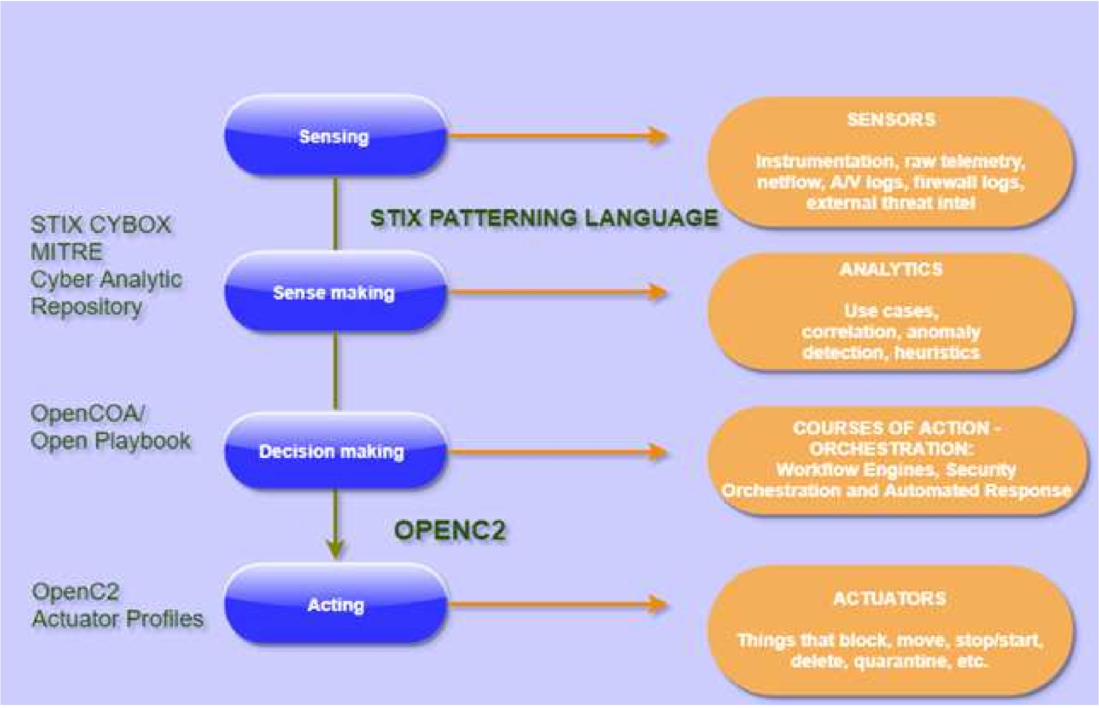

Sensing
SENSORS
Instrumentation, raw telemetry, netflow, A/V logs, firewall logs, external threat intel
STIX PATTERNING LANGUAGE
STIX CYBOX
MITRE
Cyber Analytic
Repository
Sense making
ANALYTICS
Use cases, correlation, anomaly detection, heuristics
OpenCOA/
Open Playbook
Decision making
COURSES OF ACTION - ORCHESTRATION:
Workflow Engines, Security Orchestration and Automated Response
OPENC2
OpenC2
Actuator Profiles
Acting
ACTUATORS
Things that block, move, stop/start, delete, quarantine, etc.

Figure 20. Decoupled security stack

Decoupled Security Stack assumes that (i) the whole analytics process has been completed, (ii) the decision to respond has been made, and (iii) the involved entities are authorized to respond. Moreover, benefits of decoupling are firstly, the capability to facilitate integration of new technologies and secondary, the supporting of high-level effects based on device-specific use cases. A number of extensions allow for enhanced precision to the available commands:

- ACTION: What is to be done
- TARGET: What you are doing it to
- ACTUATOR: Who is performing the command
- SPECIFIER: Identifies general to specific targets or actuators
- OPTIONS: Provide additional details for the command, target, actuator
- ID: A series of symbols or numbers which uniquely identifies an object

5.10. VERIS (Vocabulary for Event Recording and Incident Sharing)

The *Vocabulary for Event Recording and Incident Sharing (VERIS)* is a framework that enables to experts to identify a common language in computational forensics. The VERIS platform composed of *VERIS Schema Documentation* and explanation of Format, the *VERIS Community Database (VCDB)*, the *indicators of compromise (IoCB)* and the *VERIS Grid*[11].

5.10.1. Format of VERIS

[11] http://veriscommunity.net/a4grid.html

- QUESTION TEXT: Suggested wording for questions with a VERIS-based application.
- USER NOTES: Offers helpful information or tips for users of a VERIS-based application.
- QUESTION TYPE: Identifies the type of question/answer (e.g., text field vs enumerated list).
- VARIABLE NAME: Identifies the name of the schema variables.
- ENUMERATIONS: Identifies the name of the enumerates list associated with the variable. Enumerations can be found in the verisc.lib.xml or veris- enum.json documents.
- PURPOSE: Explains why we think the elements is worth having in VERIS.
- DEVELOPERS NOTES: Offers helpful information or tips to developers of VERIS-based applications.
- MISCELLANEOUS: Like it sounds; a catch- all for anything else.

5.10.2. VERIS Community Database (VCDB)

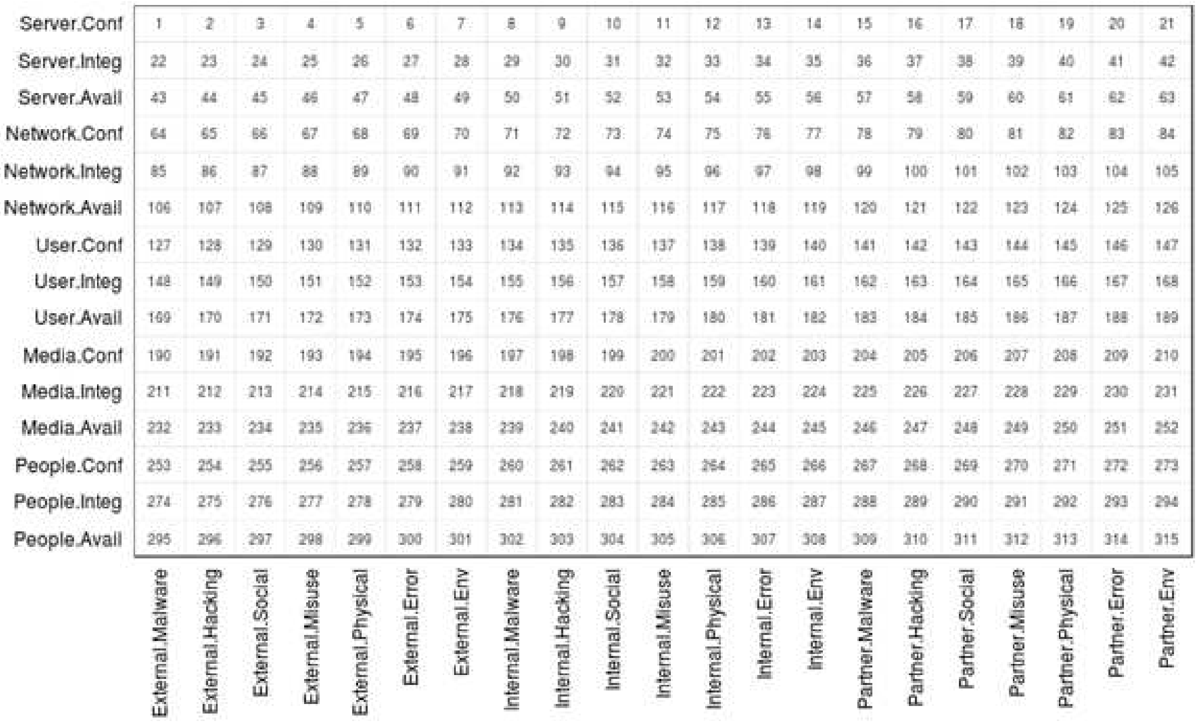
The data handling is a demand process; hence all the data locating in the base could not be published freely. So, *VERIS* Risk Team continues to drive the publication of the Verizon Data Breach Investigations Report (DBIR) annually, where they have new data-sharing partners, and they commit to keeping the report publicity available and free to download. Because the data is restricted the way someone could get them is asking *the VERIS Community Database (VCDB)* what kind of information you would like to gather. The link of Database on the GitHub https://github.com/vz-risk/VCDB.

5.10.3. Indicators of compromise (IoCs)

This section capture IoCs associated with this incident. Since VERIS focuses on strategic and risk-based information, tactical intelligence bits like IoCs are not included within the base schema. This allows them to be shared and later exported to a more suitable schema like (e.g., STIX).

5.10.4. VERIS Grid

The grid is a way to organize and visualize the main categories of actors, actions, assets and attributes in the VERIS threat model. The name derived by the 4 A-categories (3 Actos, 7 Actions, 5 Assets and 3 Attributes), if we calculate all the combinations of 4 categories we have 315 possibilities emerge (intersections).



Server.Conf	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Server.Integ	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42
Server.Avail	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63
Network.Conf	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84
Network.Integ	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105
Network.Avail	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126
User.Conf	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147
User.Integ	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168
User.Avail	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189
Media.Conf	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210
Media.Integ	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231
Media.Avail	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252
People.Conf	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273
People.Integ	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294
People.Avail	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315
	External.Malware	External.Hacking	External.Social	External.Misuse	External.Physical	External.Error	External.Env	Internal.Malware	Internal.Hacking	Internal.Social	Internal.Misuse	Internal.Physical	Internal.Error	Internal.Env	Partner.Malware	Partner.Hacking	Partner.Social	Partner.Misuse	Partner.Physical	Partner.Error	Partner.Env

Figure 21. VERIS grid[12]

[12] http://veriscommunity.net/a4grid.html

6. CTI platforms and tools

The need of assessment, detection and gathering cyber-threat information escalated over the years; this is also demonstrated by the ENISA top 15 cyber trend threats of 2016-2017, documented in [83]. More specific, the survey of ENISA indicates that the trends in cyber threats that were increased in 2016 were 9 out of 15, whereas in the year of 2017 were 11 out of 15, that means 13% increased inclination of cyber threats in one year. To address the increasing CTI needs, the frameworks described in 5. CTI formats and languages were realized into functional platforms. In this section we outline six platforms and tools (namely MISP, GOSINT, OpenTPX, Yeti, OpenTAXII, and CIF) that implement the aforementioned frameworks and language platforms for CTI sharing, also used by ENISA.

6.1. MISP - Open source threat Intelligence & Open standards for CTI sharing

One of the most widespread CTI sharing platforms is MISP-Open source threat Intelligence & Open standards for threat information sharing[13]. Following the example of most CTI sharing platforms, MISP detects, stores and shares technical and non-technical information about malware samples, incidents, attackers and intelligence. Moreover, MISP supports data export in STIX and OpenIoC for use in IDSs or SIEMSs as shown in Figure 22.

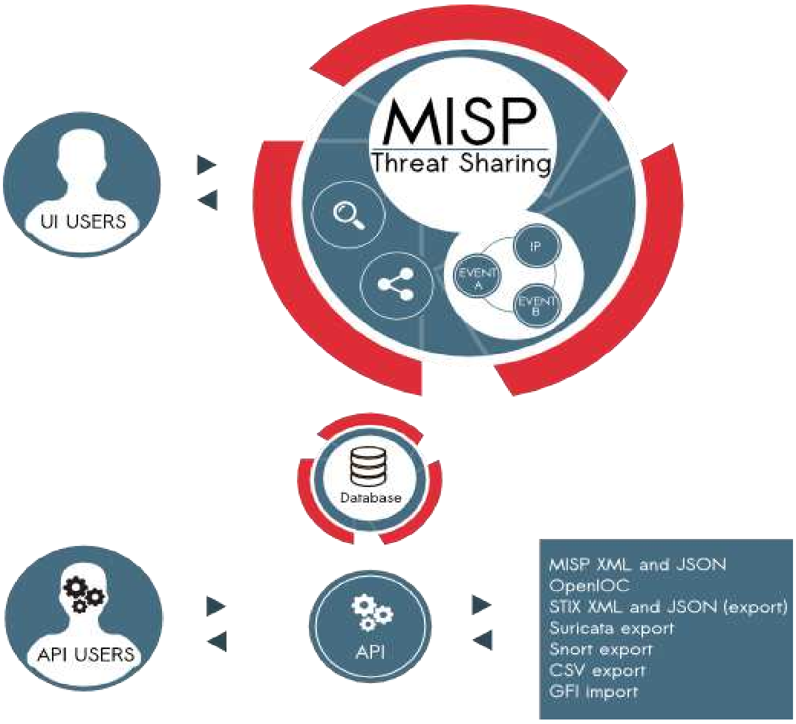


Figure 22. MISP support of standards[14]

[13] http://www.misp-project.org/

[14] https://www.circl.lu

Additionally, MISP has an automatic correlation mechanism that is: (a) able to identify relationships between attributes/objects and indicators from malware correlation engines and (b) capable of performing advanced correlations such as fuzzy hashing (e.g., ssdeep) or CIDR block matching. Another interesting characteristic of the MISP platform is that most of the supported data models are created by MISP community. MISP stores data in a structured format (to allow for the automated use of its database for various purposes), provides extensive support of cyber-security (including fraud) indicators for different vertical sectors (e.g., financial sectors), and supports CTI sharing for both human and machine applications. Furthermore, it provides STIX support, allowing data export in STIX 1.0 and 2.0 (XML and JSON) format. More details about MISP functionalities are described in https://github.com/MISP/MISP.

Intelligence vocabularies (MISP galaxy) can be bundled with existing threat actors, malware and ransomware or linked to events from MITRE ATT&CK knowledge base. MISP objects are used in the recent MISP version (2.4.80) and can be also utilized by other information sharing tools. The creation of these objects and their associated attributes is based on real cyber-security use-cases and existing practices in information sharing, while object sharing is transparently supported even for MISP instances that don't have the object template.

The MISP objects derived from many categories depend on the threat type; the supported attacks include: ail-leak (analysis information leak framework), ais-info (automated indicator sharing), android permission, av-signature (antivirus detection signature), bank account, cap-alert (common alerting protocol alert object), and others.

Finally, MISP provides a flexible free text import tool to facilitate the integration of unstructured reports into MISP and an adjustable taxonomy to classify and tag events according to the users' own classification schemes/taxonomies. The taxonomy can be either local or shareable among different MISP instances, while MISP comes with a default set of well-known taxonomies and classification schemes to support standard classification as used by ENISA, Europol, DHS, CSIRTs or other organizations.

6.2. GOSINT - The open source intelligence gathering and processing framework

GOSINT is another popular open source platform that focuses in intelligence gathering and processing. It collects, processes and exports IoCs; in this way it controls the data inclusion process in the platform and enriches it with high-quality metadata. GOSINT aggregates, validates, and sanitizes indicators for consumption by other tools including CRITs (see 7.2.1. CRITs by MITRE) and MISP (see 6.1. MISP - Open source threat Intelligence & Open standards for CTI sharing), or directly into log management systems and SIEMs, while also supporting STIX 2.0/1.x, TAXII and VERIS.

GOSINT allows forensic experts to gather structured and unstructured data from incidents occurring at third parties[15]. It is developed by Cisco CSIRT and can act as a powerful aggregator of IoCs before they are passed to another analysis platform or a SIEM. GOSINT supports also IODEF and IDMEF alongside STIX/TAXII and VERIS.

GOSINT can support several actions to provide additional context to indicators in the pre-processing phase; such actions may include the identification of IoCs with systems like Cisco Umbrella, ThreatCrowd, and VirusTotal. The information returned from these services can help analysts reach a verdict on the value of

[15] http://gosint.readthedocs.io/en/latest/

the indicator, as well as tag the indicator with additional context that might be used later in the analysis pipeline.

The GOSINT functionalities are described in https://github.com/ciscocsirt/GOSINT; the framework is written in Go with a JavaScript frontend**.** Drawbacks of the GOSINT platform are mainly related to package management and include: package managers that (a) provide out-of-date versions of the software and should be tested to ensure compatibility and (b) name packages differently depending on the package managers or OS release repository at hand.

6.3. OpenTPX - Open Threat Partner Exchange

OpenTPX[16] is a JSON (JavaScript Object Notation) based on a data model repository platform that enables to everyone hold and share incident information. It supports the sharing of several well-known protocols including HTTP, SMTP, Syslog, SNMP, FTP and others. It was created to build highly scalable machine-readable threat intelligence, analysis and network security products that exchange data at large volumes and at high speed. OpenTPX focuses on the complete picture of security and threat intelligence by providing mechanisms to convey network topology information, network ownership, network segmentation, threat metadata, threat intelligence and mitigation actions in a single mechanism. OpenTPX defines a comprehensive model of Internet associated threats that enables interconnected systems to exchange threat intelligence, context, collections, networks and mitigation information. OpenTPX was primarily designed as an optimized mechanism for data exchange at large volumes, high scale and high-speed ingestion for a broader set of Internet intelligence and threat context. Aspects of data available in STIX (e.g., indicators) have direct mapping to OpenTPX. OpenTPX operating through Threat Score Conceptual Model.

The Threat Score Conceptual Model refers to the scoring of the security landscape; it consists of 4 layers, the first layer being the core of the landscape and the fourth the surface. Specifically:

- The **1st layer** defines the network devices, topology, routing, endpoints, servers.
- The **2nd layer** defines the applications and services that run over the core layer devices.
- The **3rd layer** defines the users that run those applications
- The **4th layer** defines the observables and metadata associated with all of the 3 core layers

Risk scoring refers to all threat elements and is associated both with negative and positive observables. Each observable can be assigned to multiple classifications and each classification has an associated score. These classifications are building blocks and each block provides a comprehensive scoring framework and a query language context.

6.4. Yeti - open distributed machine and analyst-friendly CTI repository

Another open source platform is Yeti; an open, distributed, machine- and analyst-friendly threat intelligence repository[17]. Yeti is a platform meant to organize observables, IoCs, TTPs, and threat intelligence in a single, unified repository. Moreover, Yeti automatically enriches observables (e.g., by

[16] https://opentpx.org/

[17] https://yeti-platform.github.io/

resolving domains and geolocating IPs) on behalf of the user and provides a (Bootstrap-based) user interface for humans and an API-based for machines so that to facilitate communication and interoperability with other CTI tools[18]. The Yeti functionalities are described in https://github.com/yeti-platform/yeti.

6.5. OpenTAXII - Trusted Automated eXchange of Indicator Information

The OpenTAXII platform is an upgraded form of TAXII Services; its architecture follows the TAXII specifications with functional units for the TAXII Transfer Unit, the TAXII Message Handler, and other back-end services. OpenTAXII is a robust Python[19]implementation of TAXII Services that delivers a rich feature set. It provides extendable persistence and authentication layers (both via a dedicated API) and provides a collection of threat specifications. Furthermore, it provides an appropriate set of services and message exchange functionality to facilitate CTI sharing between parties. Some other characteristics of OpenTAXII include: customizable APIs, authentication, flexible logging. Furthermore, it automatically handles the data of the frameworks relied on, provides machine-readable threat intelligence, and combines network security operations data with threat intelligence, analysis and scoring data in an optimized manner. it is a large repository that consists of (meta)data of intrusions; database handling typically occurs in the same query context.

6.6. CIF - Collective Intelligence Framework

CIF (Collective Intelligence Framework) is a CTI management system and one of the platforms of choice of ENISA for CTI sharing. CIF helps users to parse, normalize, store, post-process, query, share and produce CTI data, while allowing them to combine known malicious threat information from many sources and utilize that information for identification (incident response), detection (IDS) and mitigation (null route). It also supports an automated form of the most common types of threat intelligence warehoused in CIF which are IP addresses and URLs that are observed to be related to malicious activity. The CIF framework aggregates various data-observations from different sources. When a user query for CTI data, the system returns a series of chronologically ordered messages; users are then able to make decisions by examining the returned results (e.g., series of observations about a particular actor) in a way similar to examining an email threat. The CIF Server consists of a few different modules including cif-smrt, cif-worker, cif-starman, cif-router, and ElasticSearch. The cif-smrt module has two primary capabilities: To fetch files using http(s) to/ from the local file system and to parse files using built-in parsers for regex, JSON, XML, RSS, HTML and plain text files. Moreover, the cif-worker module helps the CIF extract additional intelligence from collected threat data, the cif-starman module offers an HTTP API environment-thecif-router module provides zmq broker, while the ElasticSearch module is a data Warehouse for storing (meta)data for intrusions.

[18] *http://gosint.readthedocs.io/en/latest/*

[19] http://www.opentaxii.org/en/stable/

6.7. General and technical attributes of platforms

First, we extract from the above sections and present some key characteristics of the documented sharing platforms in the table below to allow for direct comparison.

Table 27. CTI platforms' key characteristics

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Indicative Characteristics	● Collecting ● Exchanging ● Making correlations ● Importing ● Exporting	● Collecting ● Processing ● Exporting high IoCs intrusions	● Repository ● Sharing, ● Identifying (servers-individuals)	● Organizing (objects, knowledge) ● Repository ● Managing	● Repository ● Managing ● Exchanging (services / messages) ● Exchange and mimic data and metadata	● Managing ● Sharing ● Importing ● Exporting
Objects attributes/ observables	● MISP attribute type	● N/A	● Loosely definition of threat observable	● Class of observables (Nodes)	● N/A	● Managing objects / business processes ops (CIF-Feeds) ● Command line tool to query the for observables, to generate data feeds and to submit data.
Format (Schema)	● JSON	● Go based on JSON providing by STIX	● JSON	● Python based on JSON	● Python based on JSON providing by STIX	● JSON
Github	Yes	Yes	Yes	Yes	Yes	Yes

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
How platforms face Actors	• Clusters are big objects handled by MISP GALAXY	• N/A	• Recognize pattern having Sources of threat intelligence	• Yeti is a proof of TAXII that supports the Inbox, Poll and Discovery services defined by the TAXII Services Specification	• Recognize pattern having Sources of Threat Intelligence	• Recognize pattern having Sources of CTI • The CIF SDK for Python contains library code to allow devs build applications using CIF
Formats /documentations related to other sections	• STIX • TAXII • OPENIoC • Others	• STIX • OPENIoC • CYBOX • OPENIoC • IODEF • IDMEF	• STIX	• STIX • TAXII	• CAPEC • IODEF • IDMEF • MAEC • OPENC2 • STIX 2.0 • TAXII • VERIS	• STIX

6.7.1. CTI platforms’ compliance

In addition to the above, the following table illustrates the ability of the CTI sharing platforms to meet the characteristics of requirements 3, 4 that were defined in 3. Description of methodology.

Table 28. CTI platforms’ compliance against requirements

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Interoperable	• IDS • SIEM • STIX • OpenIoC • MISPs • HTPP API • HTPP • TAXII	• CRITs • MISP • Directly into log mgmt. systems • SIEM • STIX 2.0/1.x • TAXII • OPENIoC • VERIS	• CTI frameworks • Data sources • Network sec. products • HTPP • SMTP • Syslog • SNMP • FTP	• HTTP • HTTP API • CTI frameworks • Data sources • STIX • TAXII	• TAXII Transfer Unit (TTU), • TAXII Message Handler (TMH), • Back-end services. • HTPP API	• Sources and frameworks of CTI • STIX • HTPP API • HTTP

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Expressiveness	● Intelligence vocabularies (MISP galaxy) ● MITRE ATT&CK ● Draft documents ● Training materials ● Gitter chat ● Taxonomies	● Draft documents, specifications ● Lexicons or other ● Artifacts	● Network topology information ● Network ownership ● Network segmentation ● Threat metadata, ● Threat intelligence ● Mitigation actions	● Unified repository ● Web API to automate queries	● Repository of treats	● Intelligence vocabularies (CIF-feeds)
Flexibility	● Free text import tool ● Adjustable taxonomy ● Own classification schemes existing taxonomies ● The taxonomy can be local to your own MISP	● IODEF ● IDMEF	● Comprehensive model of threat associated ● Global Internet enabling interconnected for exchanging CTI ● Threat context, collections ● Networks ● Threat Mitigation information	● Wide array of different sources ● MISP instances, ● Malware trackers, ● XML feeds ● JSON feeds	● Big range of data and a big range of data references	● Applies data that gathers from threats in repository as feeds for further analysis
Extensibility	● MISP objects are used in recent MISP version (2.4.80) ● MISP objects are in addition to MISP attributes to allow advanced combination	● External sources ● URLs ● APIs/Adhoc ● External text	● The threat observables can be extend as complex as needed	● External sources	● Persistence layer (extendable Persistence API) ● Authentication layer (extendable Authentication API)	● External sources

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Automation	● Automation supported ● New data added MISP correlated with other observables and indicators	● Automation supported ● Ad Hoc Input option.	● Machine Readable CTI ● For machine-to-machine automation, through OpenTPX JSON structure so that the TIG (CTI Gateway) can easily process the data	● Capacity to automatically enrich observables (e.g., resolve domains, geolocate IPs) ● Automation supported ● Automation to queries	● OpenTPX ● Handle use in automate way the data of the frameworks relied on. ● Automation supported	● Forms of automated reporting and mitigation services ● Automation supported
Human/ machine readable	● Both Sharing in humans and machines is supported	● Supported	● Only Machine readable	● Provides an interface for humans (shiny Bootstrap-based UI) and one for machines (web API).	● Provides machine-readable threat intelligence, in an optimized manner	● It is human /machine readable

The above tables indicate technical and general characteristics of the six referred platforms, in order to be compared. An observed outcome reached out from the comparison were the utilization of the same framework. The common threat intelligence framework is STIX. STIX and TAXII are two of the most used sources in the threat intelligence platforms. Moreover, significant and useful characteristics for selecting a platform are referred below:

- Threat feeds themselves are not intelligent. Applying contextual details and tools make them efficient so must be prioritized where possible.
- Capability to interact automatically, this capability gained by HTTP APIs. Every platform has HTPP APIs that could automatically connect with providing sources.

MISP platform

Benefits:

- Organized repository: MISP GALAXY (big objects / complex data), taxonomies, MITRE ATT&CK
- Organized community for chatting: Gitter Community
- Helpful documentation: Draft documents, Training material
- Opportunity for added tools: Free text import tool
- Creativity: Already applied and suggested data models, opportunity to make your own models

- Irrelevant information: organized GitHub, full information in internet

Highlights: The MISP platform is fully organized and the range of individuals that could utilize it could be developers or even simple users, providing materials for stand-alone learning, is very flexible and extended, automation supported. The information in the database can be extended by external sources while its functionality can be extended by integrating with third-party tools; It Is both human and machine readable. Making correlations between observables and attributes. Exceptional characteristic consisted by series of data models created by MISP community.
Drawbacks: -

GOSINT platform

Benefits:

- Organized repository: taxonomies, alert data, intrusions
- Helpful documentation: Draft documents, Lexicons, Artifacts
- Organized community for chatting: GOSINT community

Highlights: Gosint has an organized repository a managing system and exporting data. Also can be extend as source (database) by external sources (URL, TEXT, ADHOC). Platform has community that applying research from third parties to your event data to identify similar, or identical, indicators of malicious behavior, automation supported. It is both human and machine readable.
Drawbacks: Not providing updates up to date versions of the software. Package managers may name packages differently depending on the specific package manager.

OpenTPX platform

Benefits:

- Organized repository: network topology information, network ownership, network segmentation, threat metadata, threat intelligence, mitigation actions, Network security products
- Organized community: Threat Scoring framework
- Creativity: Comprehensive model of threat associated, tools of optimizing threats

Highlights: OpenTPX has an organized repository, is very flexible and extended, automation supported, we can extend the threat observables and make them as complex as you want for this reason we define them as loose (extension of data capabilities). provides a comprehensive threat-scoring framework that allows security analysts, threat researchers, network security operations and incident responders to make relevant threat mitigation decisions straight forward, while efficiently automating those decisions (threat scoring framework).
Drawbacks: it is not human readable only machine.

YETI platform

Benefits:

- Organized repository: XMLfeeds, JSONfeeds, taxonomies
- Organized community: Yeti GitHub Community
- Helpful documentation: MISP instances

Highlights: Yeti has an organized repository, is very flexible and extended, automation supported. Extended as source(database). It Is both human and machine readable. Yeti, goals is to turn it into a self-sustainable project, where not only the core developers but the whole community helps out when the community

needs help (they don't have achieve it yet). The communication handled centrally for this reason only in GitHub, GitHub issues for all communication.
Drawbacks: Do not provide tools for creation of incident attack. Do not making correlations between observable and attributes.

OPENTAXII

Benefits:

- Organized Repository: references, data and metadata of threats, mitigation actions

Highlights: It is assumed that OPENTAXII has an organized repository and managing system, also can mimic already known cases and threats. It is flexible and extendable since it is providing machine-readable threat intelligence, possibility of layer extension, source intelligent extension and APIs extension Provides automation.
Drawbacks: Does not provide tools for creation of incident attack.

CIF

Benefits:

- Organized repository: CIF intelligence vocabularies (CIF-feeds), Combination of malicious threats

Highlights: CIF has an organized repository and managing system and exporting data. Provides combination of malicious threats and utilize that information for identification (incident response), detection (IDS) and mitigation (null route). Extended as source (database) indicators of malicious behavior, automation supported, human machine readable. It Is both human and machine readable.
Drawbacks: Only observed threats (such as IPs).

6.7.2. CTI platforms' scoring

Based on the above discussion, we can proceed to map in simple manner the extent to which properties of requirements 3 and 4 are being met by the candidate platforms. Requirement 6 is not depicted here as all the tools considered are already chosen to be open source. A score value of '1' (resp. '2') suggests that the specific property is present to a satisfying (resp. high) level, or '—' to indicate that it is not.

Table 29. CTI platforms' scoring

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Interoperable	2	2	1	1	2	1
Expressiveness	2	1	2	1	1	1
Flexibility	2	1	2	2	1	1
Extensibility	2	2	1	1	2	1
Automation	2	2	2	2	2	2
Human/machine readable	2	2	—	2	1	2

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Overall score	12	10		9	9	8

All platforms support a number of standards from those presented in 5. CTI formats and languages (as shown in Table 28 and Table 29), where STIX is the bottom line. It is clear from the above comparison that MISP and GOSINT are taking the lead in platforms' race; the final decision will be made in 8. Recommendations after having also considered the findings of 7. Current market situation.

7. Current market situation

Organizations worldwide, from governments to public and corporate enterprises, are under constant threat by evolving cyber-attacks. The fact that there are literally billions of IoT devices globally, most of which are readily accessible (via Telnet) and easily hacked (due to lack of security controls), allows threat actors to use them as the cyber-weapon delivery system of choice in many today's cyber-attacks, e.g., from botnet-building for launching distributed denial of service attacks, to malware spreading and spamming. The sooner an organization knows about emerging threats, the more efficiently cyber-defense mechanisms will be utilized. Therefore, the main challenge organizations face is the abundance of data and the lack of actionable intelligence.

This section reviews several market solutions related to the discovery and management of CTI, and present their main features and characteristics with respect to a number of different facets including architecture, offered services, standards' adoption, and mode of operation. The solutions' presentation is organized in the following areas:

1. **Human intelligence services.** Services delivering intelligence reports, which are the result of human analysts' work on an incident, that present the techniques, tactics, and procedures (TTPs) used by threat actors. In some cases, advanced analytics and machine-learning techniques are used to increase the efficiency of analysts and produce faster and more accurate reports.
2. **Threat data feed providers.** These are organizations that provide (possibly aggregated) data feeds on potential threat indicators, such as IP addresses, domain names, and file hashes, that users can subscribe to. Often, there is a lack of context in these indicators (e.g., links to specific technologies, attacks, etc., or correlations between the collected threat information).
3. **Threat intelligence platforms.** These help organize (thousands of) threat data feeds into a single container and manage them in a centralized way, e.g., by removing duplicate entries, prioritizing the sources of data, and configuring alerts, as well as to integrate them with other security products (e.g., SIEMs) or incident response solutions.
4. **Complete threat intelligence solutions.** A complete solution that combines the capabilities of the above (services, providers, and platforms), in order to allow getting the most out of the available CTI in nearly real-time.

Social, mobile, digital and collaboration platforms, such as Slack, LinkedIn, Facebook, mobile app stores, Pastebin, Microsoft Teams and more, have become the new cyber-security battleground, presenting one of the largest and most dynamic risks to organizational security in decades. In the social media age, exploiting people to extort sensitive information or gain access to critical systems has never been easier. Targeted cyber-attacks can now hit an organization's employee via social media. As a result, there is a growing trend in cyber-security products to incorporate CTI being the result of social media analysis; the same holds for the wealth of information that exists in the deep/dark web. This is mostly evident in the description of the CTI solutions included in the first area (*see* the next subsection).

7.1. Threat intelligence services

7.1.1. BreachAlert by RepKnight

RepKnight's BreachAlert engine (https://www.repknight.com) provides automatic monitoring of the dark web and early warning services if your data, sites, applications or assets are being discussed in these forums. The monitoring includes places like the Tor network, chatrooms, and text dump and bin sites (like Pastebin and Ghostbin) that are not indexable by search engines (like Google and Bing). RepKnight also monitors places that require human intervention (e.g., require login, invitation or contain CAPTCHA) through a combination of in-house analysts and external covert sources. These places are used by hackers and cyber-criminals to discuss and share information about vulnerabilities and exploits.

RepKnight's data collectors are constantly scraping data from the monitored places collecting millions of posts every day and copy the textual information residing on the monitored places into RepKnight's database. Following, they continuously search for information pertaining to possible attacks. In such cases, RepKnight alerts the organization to anything new matching its data. For instance, any mentions of the company sites, products or IP addresses in the monitored places possibly indicate that an organization has been targeted by cyber-criminals or even that it has already been hacked.

In addition, BreachAlert searches the collected data from the dark web for stolen or leaked information by incorporating watermarks and fingerprints. *Watermarks* are fake, but plausible, data records, which are inserted into a database as markers; for example, by adding a fake customer to the Customers Relationship Management system of an organization. Such watermark data should never appear outside your organization. Finding these watermarks elsewhere (e.g., in the dark web), indicates a potential breach. *Fingerprints*, on the other hand, are characteristics of data which uniquely identify them. For example, it might be the format of an account number, or a list of the home postcodes of an organization's employees. If such information is found in a single dark web post, it may indicate a breach of the Human Resources database. Fingerprints do not require the source data to be modified and can be used to search for data which may not be practical to watermark (for instance, payroll instructions, or online credit card transactions). By design, both watermarks and fingerprints do not contain sensitive information or personally identifiable information. Thus, they can be safely used outside of an organization, for example as search terms in BreachAlert.

Overall, RepKnight's BreachAlert platform acts like a burglar alarm for an organization's data, providing early warning of potential breaches, and demonstrating proactive mitigating steps to minimize the effects of any breach.

7.1.2. DyTA by Cytegic

The Cytegic Dynamic Trend Analysis (http://www.cytegic.com) solution is a novel intelligence analytics platform that enables cyber-intelligence analysts to identify cyber-threats related to geopolitical regions and business sectors, based on data from online open sources and technical cyber-feeds. DyTA automates the gathering, processing and analysis of data, and generates specific and actionable cyber-threat forecasts, based on built-in pattern analysis capabilities.

DyTA technology enables managers and cyber-analysts to analyze threats efficiently and generate actionable intelligence to prepare for a wide range of cyber-attacks. Additionally, DyTA forecasts potential cyber-threat trends relevant to an organization. DyTA is formed by the following modules:

- **Data collection.** Gathers mass volumes of data from open-sources including blogs, forums, websites, and technical cyber feeds.
- **Semantic analysis.** Automates the analysis and synthesis of entities for any geopolitical and business sector context.
- **Behavior profiling.** Generates contextual intelligence about attacker activity patterns during the entire event life cycle.

- **Forecasting.** Utilizes statistical and behavioral analysis tools to prepare against potential cyber-threats.

Overall, the security experts are allowed to (a) know their adversaries by identifying the objectives and capabilities of attackers relevant to an organization, (b) measure the corresponding threat level by quantifying threats by different adversaries to specific business assets, (c) increase readiness by utilizing advanced forecasting methods to prepare organizations for potential cyber-threats and (d) enhance collaboration by exporting CTI to internal and external systems to increase awareness and preparation.

7.1.3. F5 Labs

The hunt for exploitable IoT devices by attackers, has been a primary area of research for F5 Labs (https://f5.com/). More specifically, F5 Labs monitors the dark web for the latest malware variants and threat actor behaviors, collects global attack data, and creates threat monitoring tools. F5 Labs publishes the team's intelligence including information about which malware is targeting whom, how the malware works, what attack trends the team is seeing.

F5 Labs processes the collected information with experienced security researchers to provide actionable intelligence on current cyber threats and to identify future trends. F5 Labs looks at everything from threat actors, to the nature and source of attacks, to post-attack analysis of significant incidents to create a comprehensive view of the threat landscape.

F5 Labs protects customers from malware, phishing, and web fraud with proactive, 24/7 real-time global threat monitoring. F5 Labs threat intelligence leverages the SOC's real-time analysis of current threat actors and attack trends. F5 Labs aims to protect organizations from the newest malware variants and zero-day exploits and attack trends by exploiting the latest insights gathered from the F5's threat intelligence team.

7.1.4. Flashpoint Intelligence Platform

Security startup Flashpoint provides Business Risk Intelligence (BRI) to its customers and partners. The Flashpoint Intelligence Platform (https://www.flashpoint-intel.com/solutions/) gains access to deep and dark web communities, to allow users investigate them safely and supplement their internal data with targeted information from highly curated sources and subsequently gain greater context around any information they might need in order to draw their own conclusions. Through the platform, users also gain access to the Finished Intelligence dataset, which contains analytical reports and original primary source data used by experts to create those reports. Those intelligence reports cover a wide spectrum of illicit, underground activity, from cybercrime and hacking, to fraud, emergent malware, DDOS intelligence, hacktivism, violent extremism, and physical threats. This, not only allows organizations to save time and resources in monitoring the deep and dark web, but also informs and protects them without any additional risk. The platform offers also a Risk Intelligence Observables (RIOs) dataset, which provides a high-fidelity feed of cyber observables that customers can integrate into their security operations to give them decision advantage. RIOs enrich user data with additional context, empowering customers to better understand and mitigate risks.

Finally, Flashpoint's API grants users the ability to monitor and set up alerts for the use of certain keywords to help with specific threats or risks. Due to the types of data that Flashpoint is exposing in the API, the format and structuring that STIX/TAXII provides is not suitable for information sharing, so a JSON-based approach is used instead[20].

7.1.5. Intel 471

Intel 471 (https://intel471.com/) provides an actor-centric cyber intelligence collection capability. This collection focuses on infiltrating and maintaining access to closed sources in both dark and surface web, where threat actors collaborate, communicate and plan cyber-attacks. Such sources contain places where entry is highly guarded; for instance, underground marketplaces or chat rooms. Following the actor-centric approach of the cyber intelligence collection, Intel 471 helps companies identify who is responsible and what is the motivation of an attack. This fact allows companies to become proactive by unearthing the likely attackers in advance of a perceived threat. In their knowledge base, Intel 471 houses over 10 million actor handles and linked data, tracking them across various marketplaces and providing deep contextual insight about imminent threats.

Due to this innovative approach, Intel 471 delivers unmatched visibility into actors and their TTPs, planning, marketplaces and communication networks; information useful for the existing security systems to expand their knowledge base, adding two more parameters to the cyber threat data collected. For that reason, various CTI systems have adapted to this approach by integrating with Intel 471. Amongst them are the ThreatStream [32], ThreatConnect[21], ThreatQuotient[22], EclecticIQ [31] and Recorded Future[23].

Intel 471 delivers two products[24] differentiated by actor motivation:

- **CyberCrime.** This product covers cyber threat activity from financially motivated cyber criminals. It allows for better understanding of the latest tools and techniques available in the cybercriminal underground and targets the identity of the actors, their motivation and how they fit into the cybercrime community.
- **Hacktivism.** This product covers cyber threat activity from politically motivated cyber criminals. It allows the better understanding of current targeting, leadership trends, hacktivist groups and movements, motivations, and the TTPs related to this threat activity.

Along with the above products, Intel 471 provides access to the online portal that supports full text searching, monitoring actors across forums and social network analysis. Moreover, Intel 471 provides an API which allows automated queries by handle, IP addresses, and other. Thus, users can conduct detailed searches and set up watchers with varying frequency to track the activity of actors or other information that Intel 471 follows. Finally, Intel 471 provides access to third party TIP integrations.

7.1.6. LookingGlass Threat Intelligence

LookingGlass Cyber Solutions[25] has three distinct SaaS packages: (a) Information SaaS, (b) Brand SaaS and (c) Physical SaaS. Each package delivers continuous monitoring of the Internet (surface, social, deep and

[20] http://www.eweek.com/security/flashpoint-digs-into-dark-web-with-security-intelligence-api

[21] https://www.threatconnect.com/partner-type/threat-intelligence/

[22] https://www.threatq.com/integrations/

[23] https://www.recordedfuture.com/partner-spotlight-intel-471/

[24] https://intel471.com/#products

dark web) for real-time indicators of threats. Customers also receive granular incident notifications delivered via portal or API, as well as a monthly summary of incidents.

With the Information SaaS package, in particular, the analysts and systems identify information security threats both on the company's and on open source systems. Examples of information security threats observed on company systems are malware hosting and distribution, virus and botnet infection, command-and-control activity, malicious and scanning behavior, hosting phishing activity, ransomware and sending spam. Examples of information security threats observed on open source are compromised company account credentials, suspicious domain registration, phishing and spoof sites, sensitive, confidential, IUO document and information disclosure, search for sensitive data, discussion by known threat actors, discussions and threats (in the dark web, hacker forums, IRC, etc.) and data breach announcements.

7.1.7. Norse Intelligence

Norse Corporation Solutions [50] are built upon the Norse Intelligence Network, which is a globally distributed grid of millions of sensors, honeypots, crawlers, and agents that provide visibility into the surface and dark web, where bad actors operate. Processing hundreds of terabytes daily, encompassing social media, categorization, malware augmentation, 5 million to 20 million emails, whois and whowas for all domains seen, URL/IP context and threat taxonomy, Norse Intelligence Network computes over 1500 unique risk factors for millions of IP addresses. Consequently, the Norse network continuously analyzes traffic and provides insight to identify the compromised hosts, malicious botnets, anonymous proxies, and several other sources of attack.

The devices that constitute the Norse Intelligence Network, are the Norse Appliance [46, 47] machines. Operating 24/7 on the grid, Norse Appliance supplies its users with pre-filtered and pre-processed logs with full context, deriving from the global network's analysis of all data sources that are being monitored. Currently, Norse vastly provides CTI solutions to the government sector via the integration of Norse Appliances with their security systems, aiming to provide their solutions in the financial and technology sectors in the future as well. Finally, users can easily configure the Norse Appliance devices in their organizations, either through the integrated hardware (LCD touchscreen), or via virtual drivers.

Finally, on top of the Norse Intelligence Network, Norse corporation developed the Norse Intelligence Service [49], which provides a visualization of Norse network's collected data. Through the utilization of Norse Intelligence Service, users can view the continuously monitored cyber threat sources with local, supply chain and global relevance. In addition to that, Norse Intelligence Service portal delivers monthly reports about actionable threat trends and actor analysis, which indicate compromise or direct targeting, as well as recommended actions to mitigate the threats.

7.1.8. SearchLight by Digital Shadows

Digital Shadows SearchLight[26] combines scalable data analytics with human data security experts to provide a holistic view of an organization's digital risk profile including cyber threats, data and brand exposure, VIP exposure, infrastructure exposure, physical threat and third-party risk. The process of the analysis includes:

[25] https://www.lookingglasscyber.com/products/assess-risk/threat-intelligence-service/

[26] https://www.digitalshadows.com/products/digital-shadows-searchlight/

- **Planning and collection.** SearchLight continuously monitors the open, deep, and dark web for mentions of a company's assets and unique identifiers. Using a diverse range of collection techniques, SearchLight monitors across the broadest range of data sources, including paste sites, code repositories, mobile app stores, technical forums, file hosting and sharing sites, messaging platforms, criminal forums and marketplaces, news sites, threat actor blogs and mouthpiece sites, security research blogs, infrastructure search engines, exploit and vulnerabilities libraries, social media profiles, groups and posts and the dark web via Tor and I2P in order to collect a full picture of the organization's digital risk.
- **Automated analysis.** Irrelevant mentions are removed through a combination of data science and machine learning.
- **Human analysis.** Expert analysts verify automated incidents, greatly reduce false positives, conduct further research, add context, and assign a severity level.
- **Dissemination.** Relevant, prioritized and actionable incidents are delivered via the SearchLight portal, email alerts or API.

Digital Shadows SearchLight supports STIX/TAXII tags as elements for creating the digital risk profiles.

7.1.9. Security Ratings by BitSight

The BitSight Security Rating Platform (https://www.bitsighttech.com/security-ratings) generates objective, quantitative measurements on an organization's security performance to produce daily security ratings. Those ratings are produced by analyzing existing security incidents and practices, using externally observable, non-intrusive data and methods. Daily security ratings enable organizations to proactively identify, quantify and manage cybersecurity risk throughout their ecosystem.

The BitSight Security Ratings are calculated by collecting input from extensive and diverse sources. Such input can be communication with known botnets, spam, malicious code, whether the company has preventative security configurations, DDoS, user behaviors, news feeds, social media and more. Then, that input is processed by a proprietary algorithm which normalizes and maps the input with parameters such as type of compromised system, date, company IP, severity, frequency, duration and confidence, and then summarizes the security risk in the customer portal in the form of analytics, ratings and alerts.

The Security Ratings can be used by organizations in various ways. One such usage includes benchmarking the security performance to enable organizations quantify their cyber risk, measure the impact of risk mitigation efforts, and benchmark their performance against industry peers. The company can view their own compromised systems and diligence data, and use them to better identify the sources of risk and take swift action to mitigate it. There are also some enhanced capabilities such as BitSight Forensics, that can provide a company with specific information on communications with Command and Control servers to remediate potentially harmful infections.

Another way that an organization can take advantage of the Security Ratings, is to manage risk posed by third party. The third party in question could be vendors, potential new clients, business partners or acquisition targets, and the Security Ratings can help organizations quickly and cost-effectively mitigate risk within third party networks. With these ratings, businesses can proactively identify issues within their extended network ecosystem, prioritize and streamline further assessment and drive conversations about security controls.

7.1.10. Wapack Labs Cyber Threat Analysis Center

Wapack Labs is a company which offers cyber threat intelligence services. Their main service, Cyber Threat Analysis Center (https://www.wapacklabs.com/ctac/), is a web dashboard containing multiple cyber tools such as Elastic Stack, CyberChef, RocketJot. All these tools are aimed at helping cyber threat analysts interact with all data collected by Wapack Labs, which are derived from the monitoring of several sources such as dark web content, keylogger outputs, sinkholed connections, pastebin mentions, malicious emails detected and open source material. Through Kibana (https://www.elastic.co/products/kibana), the users are able to search and visualize the Wapack Labs Elastic Stack for information about cyber threats that might concern them, such as the cyber threat type, the malicious involved IPs, the type of data that might have been leaked, the location of a sinkhole, etc.

Finally, with the development of RiskWatch (https://www.wapacklabs.com/riskwatch/), Wapack Labs extend their solutions with pub/sub functionality. RiskWatch is a service provided to registered users who are interested in receiving a cyber risk report conducted by Wapack Labs analysts. By regularly monitoring the aforementioned sources, Wapack Labs notifies registered users about suspicious activity on a weekly or monthly basis, reports all current known and ongoing threats, and provides critical insight into an adversary's intentions and activities.

7.1.11. ZeroFox

The ZeroFOX Platform (https://www.zerofox.com) delivers automated threat detection and remediation across social, mobile, digital, and collaboration platforms. ZeroFOX identifies organizational risks and security threats targeting both businesses and employees. Using diverse data sources and artificial intelligence-driven analysis engines, the ZeroFOX Platform automatically identifies and remediates fraudulent accounts, phishing attacks, customer scams, exposed PII, insider threats and more.

ZeroFOX protects organizations from cyber, brand and physical threats on social media and digital platforms by achieving the following tasks.

- **Gain visibility and control.** Using ZeroFOX, security analysts immediately see beyond the perimeter of the organization to identify fraud, targeted attacks and risk faster, before they impact on business, customers, partners and employees.
- **Access immediate situational awareness.** Physical threats identified by ZeroFOX in real-time enable corporate security teams to be as prepared as possible, giving true visibility into dynamic social media and digital platform data for situational awareness.
- **Address the problem at scale.** ZeroFOX leverages artificial intelligence and automation to reduce time-intensive collection, analysis and remediation. Thus, it may scout trillions of potentially malicious accounts and posts.
- **Find leaked data and gain context.** ZeroFOX help security teams to pinpoint indicators of attacks being planned or leaked data after an attack broadcasted on social media and external digital platforms.

7.2. Threat data feed providers

7.2.1. CRITs by MITRE

Collaborative Research Into Threats (CRITs) (https://crits.github.io) is MITRE's open source malware and threat repository that leverages other open source software to create a unified tool for analysts and security experts engaged in threat defense. It has been in development since 2010 and aims to provide the security community with a flexible and open platform for analyzing and collaborating on threat data.

CRITs is free and open source; its source code is available from the CRITs GitHub repository (https://github.com/crits). Users of CRITs may upload threat data and uncover critical information to keep their organization safe. CRITs allows the management and the interpretation of vast quantities of intelligence data in a single repository. Moreover, CRITs users may develop additional capabilities using the services framework to combine CRITs with third-party and home-grown intelligence systems. Users of CRITs mostly use Python, JavaScript and CoffeeScript.

Therefore, CRITs provides users and organizations around the world with the capability to quickly adapt to an ever-changing threat landscape. It can be installed locally for a private isolated instance or shared among other trusted organizations as a collaborative defense mechanism.

CRITs works with developers everywhere in the globe to discover and share capabilities for threat defense. It gathers information from everyone who has an idea on how to make it better, in order to ensure that it can identify and prepare for the next big technological breakthrough. Finally, CRITs constantly improves collaborative features and services so everyone can contribute and participate in threat research and analysis.

7.2.2. Enclaves by TruSTAR

Enclaves (https://www.trustar.co/product/threat-intelligence-platform) is TruSTAR's secure data repository used for storing, managing, and enriching sensitive events. It allows users to analyze and enrich investigations with trusted, relevant intelligence sources, including information shared by their partners and peers, while allowing them to maintain protective access controls.

Enclaves was built to address the challenges posed by modern analysts. To this end, TruSTAR created a platform with their needs in mind. Some key features of TruSTAR include:

- **Analysis visualization**. Enclaves provides a human-analyst-friendly visualization. Its goal is to illustrate how an IoC interconnects with existing Enclaves threats.
- **Extensible interoperability**. Enclaves supports STIX and TAXII and contains a variety of build-in integrations in its marketplace. Additionally, Enclaves provides SDKs and a robust RESTful API to enable integrations with proprietary data sources, ticketing or case management systems, or any other SOC tools as needed.
- **Machine learning-assisted extraction**. Enclaves integrates an extraction engine capable of automatically identifying an industry-leading 12 types of IoCs from structured and unstructured data, instantly surfacing them for further analysis. Also, the platform can be programmed to support different flavors of IoC.
- **Collaboration, sharing and automated redaction**. Enclaves supports data sharing by allowing the granular definition of user permissions (who can access and interact with what data), thus, adhering to any compliance requirements. With in-app chat and the ability to capture notes on

investigations, Enclaves users and teams are empowered to collaborate with ease in their task to add context to ongoing analysis and IoCs. Additionally, Enclaves redaction engine ensures that sharing remains legal by allowing visibility control of all shared data. For instance, the engine may scrub sensitive information from reports before releasing them to public or partners. To identify potential PII to redact, Enclaves uses a natural language processing engine.

- **Custom tagging**. Enclaves uses custom tags to organize incident reports. Tags can be based on any proprietary naming schemes, such as department names, threat families, or ticket numbers. Analysts may use tag to locate information during investigations.
- **Search for context and IoCs**. Through Enclaves users may search for IoCs (by threat name, hash, IP, domain, etc.) and surface relevant context from relevant investigations and external intelligence sources. Results displayed in the analysis visualizations make it easy to pinpoint patterns, discover trends and hunt within TruSTAR.
- **Notifications and alerting**. Enclaves allows users to save their current analysis and enable notifications. By doing so, whenever another analyst adds new context to your case or other correlating IoCs or relevant cases become available, appropriate analysts are notified and may utilize this new context.

7.3. Threat intelligence platforms

7.3.1. ActiveTrust by Infoblox

ActiveTrust (https://www.infoblox.com/products/activetrust/) is an Infoblox product [33], which provides access to a platform, either by on-premises system installation or by SaaS (via ActiveTrust Cloud), that guides users to proactively detect and prevent cyber threats. To do so, ActiveTrust bundles various Infoblox tools such as Infoblox DNS Firewall, Infoblox Threat Insight in the cloud, Infoblox Threat Intelligence Data Exchange, and Infoblox Dossier. The main idea of ActiveTrust is to intercept DNS traffic, in order to counter DNS-based data exfiltration or leaks and malware communications with command and control hosts. In addition, it is an optimal approach for devices which cannot implement agent software such as PoS, medical equipment, certain IoT devices, etc.

Infoblox DNS Firewall executes only administrator-defined policy action (blocking, redirecting devices to a protected site, and/or logging events), to help stopping devices from communicating with command and control hosts or botnets via DNS.

Infoblox Threat Insight in the cloud, being offered as a cloud service, prevents (with DNS Firewall) DNS-based data leaking or exfiltrating by making use of reputation, signatures, and behavioral analytics data stored in the Infoblox knowledge base. In addition, it is capable of blocking newer threats such as DNSMessenger, DGA, and Fast Flux.

Infoblox Threat Intelligence Data Exchange leverages accurate machine-readable threat intelligence data, to correlate, combine and selectively distribute data across other security infrastructures. Infoblox threat feeds begin with information gained from harvesting and monitoring techniques and are then correlated and combined with verified and observed data from Infoblox trusted partners. Finally, the cyber threat information collected is being normalized and refined by the Infoblox threat intelligence team, in order to severely reduce false positives.

Infoblox Dossier is a threat investigation tool, which provides search capabilities (limited up to thousands queries/year) to threat context contained in the Infoblox knowledge base, in order to help threat analysts simplify their investigation methods.

Also, with the Infoblox Security Ecosystem license, users are enabled to integrate Infoblox DNS Firewall with third-party security systems like iSight by FireEye, QualysGuard Vulnerability Management and other threat intelligence platforms, sharing DNS data such as IoC, in order to initialize a scan when a new device connects to the network and determine whether it is malware infected.

Consequently, viewing into the capabilities provided by each tool separately, Infoblox ActiveTrust is a complete CTI solution which prevents DNS-based data exfiltration, detects and blocks DNSMessenger, DGA and Fast Flux, stops DNS-based malware command and control or botnet communications, and collects threat intelligence from internal and external sources into a single platform.

Finally, Infoblox ActiveTrust supports the creation of custom API data feeds which combine threat data from all monitored sources, use contextual metadata to select a relevant subset and are able to be exported in a standardized format such as JSON, STIX, CSV, etc., in order to improve users' existing security ecosystem.

7.3.2. BrightCloud Threat Intelligence Services by Webroot

BrightCloud Threat Intelligence Services (https://www.webroot.com/us/en/business/threat-intelligence) provide predictive threat intelligence on URLs, IPs, files, and mobile apps, which can protect organizations even from previously unknown attacks. These services leverage the Webroot TIP, a cloud-based security platform, which is enhanced by a contextual analysis engine to correlate information for deep insight across the online threat landscape. This advanced self-learning platform continuously scans the internet and incorporates inputs from millions of sensors, so that BrightCloud services can quickly and accurately identify previously unknown threats.

BrightCloud Threat Intelligence Services utilize a three-dimensional approach to provide threat intelligence that includes:

- Breadth of services to cover all critical threat vectors.
- Big data architecture for volume, scale, depth, and speed.
- Greater accuracy and predictive risk scoring by correlating previously disparate data derived from real-world endpoints.

Some of the services provided (however, not from a single platform) include (a) Web Classification Service, which classifies and blocks malicious and unwanted web content based on 82 categories, (b) Web Reputation Service which offers a reputation score that forecasts the security risk of visiting a website, (c) IP Reputation Service, which provides a dynamic list of 8 to 12 million malicious IPs at any given time to block malicious traffic from entering a network and can also give access to additional intelligence and leverage predictive risk scores to customize network settings based on needs and risk tolerance, (d) Real-Time Anti-Phishing Service, which provides time-of-need protection through real-time scans before sites are visited, to catch advanced phishing attacks that can expose an organization to breaches and data loss, (e) Streaming Malware Detection, which is designed to combat polymorphic malware by providing a risk score for files as they traverse the network perimeter to enable users to quickly allow, block, or flag files for investigation and the (f) Mobile Security SDK, which can be embedded into a mobile application to provide protection against mobile threats through antivirus, antimalware, device and application interrogation, secure web browsing, and web classification, along with device risk scores.

BrightCloud Threat Intelligence Services use a powerful contextual analysis engine that takes disparate data from Webroot TIP feeds and correlates it for deep insight into the landscape of interconnected URLs, IPs, files and mobile apps. Mapping the relationships between these different data points enables Webroot to provide partners with highly accurate and actionable intelligence that is always up-to-date. This also allows Webroot to accurately predict how likely an internet object is to be malicious in the future by its associations with other URLs, IPs, files, and mobile apps. For example, a seemingly benign IP, which other services may classify as safe, may be tied to other URLs, IPs, files, or mobile apps with histories of dangerous behavior. This advanced analysis provides a predictive reputation score which enables users to proactively protect themselves through self-defined policies based on their risk tolerance.

Finally, BrightCloud Threat Intelligence Services integrate with existing security solutions through the Webroot SDK and a REST API. Depending on the service, it may be integrated in three modes (hosted, local database, or hybrid), allowing partners to select the integration and deployment type best suited to their needs.

7.3.3. Cyber Advisor by SurfWatch Labs

The SurfWatch Cyber Advisor[27] solution combines analytics, products and human experts to establish CTI. The solution takes a lifecycle approach that first builds a personalized cyber risk profile. This profile is created by performing a baseline risk assessment on the network in order to discover potential vulnerabilities. Next, the cyber risk profile is continuously monitored against relevant, trending threats by collecting cyber threat data from several sources, such as cyber expert blogs and news feeds, CVEs, open source material, government data breaches, phishing reports, dark web markets and forums, and paste sites. Then, all collected data is being processed using natural language and human processing methods, in order to keep only the useful information and reduce false positives. Subsequently, if new threats appear in the spotlight, SurfWatch Cyber Advisor alerts its users by delivering a finished intelligence report based on the personalized cyber risk profile. The alerting process can be customized in terms of both the reports' recipients and the frequency of these notifications. Along with these reports, users receive a prescribed set of practices, deriving from in-depth analysis of the specific cyber threats, in order to mitigate the risk. Finally, Cyber Advisor provides updated lists of malicious URLs and IPs based on SurfWatch Labs knowledge base.

SurfWatch Threat Analyst (https://www.surfwatchlabs.com/threat-intelligence-products/threat-analyst) is a SaaS product that provides cyber threat intelligence to help organizations identify possible attacks and rapidly mitigate cyber risks. Threat Analyst automatically collects, monitors and tracks relevant threats from open and dark web sources, providing information about cyber risks of businesses, supply chain and industry. Thus, the procedure of the cyber events impact classification on key areas of businesses becomes straightforward. Hence, CTI teams can immediately create a cyber strategy to eliminate all imminent risks. Moreover, SurfWatch Threat Analyst provides visualized information about possible cyber risks that encompass a business's supply chain, which could also impact the business itself. In addition to having access to such information, Threat Analyst also provides a notifications system in order to keep businesses alarmed in case that an intelligence group is indicated for a relevant threat. Next, through the SurfWatch

[27] https://www.surfwatchlabs.com/threat-intelligence-products/cyber-advisor

Threat Analyst platform, users have access to tailored and flexible dashboards that provide a visualization of trending and relevant cyber risks. Finally, SurfWatch delivers cybersecurity news in order to keep users alerted on the latest developments and cyber events.

Along with Cyber Advisor and Threat Analyst, SurfWatch developed the Analytics API[28] in order to export cyber threat information in standardized formats. Through the SurfWatch CyberFact data model, raw cyber event information is automatically transformed into CyberFacts, which provides information about who's behind the attack, what is their target, what the effect is and what method is followed to carry out the attack. Industry target tags accompany every CyberFact and they describe which business or organization is most impacted by the event. In addition to that, the CyberFact data model helps to the creation of CyberInsights which derive from the analysis of CyberFacts and they provide insights about recent evolving risks. Finally, SurfWatch Analytics API delivers CTI in JSON format over REST by either pushing to the designated endpoint in real time or by querying. Using the API, users can get: 100 queries/minute, JSON data format, compressed data transport, instant delivery of all CyberFacts and CyberInsights, real-time data delivered over HTTP POST, queried data accessed by HTTP GET, and threat intel feed available in STIX/TAXII 2.0 formats.

7.3.4. Cyjax Technology Stack

The Cyjax Technology Stack (https://www.cyjax.com/site/threat-intelligence-platform) is a portfolio of custom-built technologies designed to provide automated collection, processing, monitoring and advanced analytical capabilities of threat intelligence information. The prominent modules of Technology Stack are:

- **Cymon user dashboard.** The Cymon user dashboard is the base of the platform. It brings all the information outputs together to provide custom views of real-time threat intelligence data, alongside an essential set of advanced tools that deliver cutting edge analytical capabilities to end users.
- **Daily brief.** The daily briefing service is a real-time incident reporting feed, available through the platform or via API. Using optional industry verticals, users can view a running commentary of the day’s key cyber events within their sector, or those around them.
- **DarkWatch.** Mapping, monitoring and mirroring the dark web’s most prolific marketplaces, forums and websites, the DarkWatch module helps the user to identify brand and personal exposure, discover emerging threats and understand the shape they are taking and how they may impact their assets in the near or long-term. Through DarkWatch users can interact with and study the dark web with no footprint and at no risk to them or their organization.
- **PasteWatch.** PasteWatch is a paste monitoring service and it is built to capture third-party data asset exposure. Collecting pastes across multiple websites, PasteWatch automatically extracts intellectual property from user credentials and exposed credit card data, to outsourced software development. The PasteWatch service can push credentials and exposed credit card numbers directly to the user's SOC upon discovery.
- **TweetWatch.** With a built-in customisable sentiment analysis dictionary, the Twitter monitoring capability allows the user to define the threat triggers that relate to their requirements. This helps

[28] https://www.surfwatchlabs.com/threat-intelligence-products/analytics-api

users to stay on top of incoming social media Tweets, identifying trends and patterns which quickly determine any emerging threats or threat actors.

- **NewsWatch.** Scanning thousands of news sites in over 100 countries, the NewsWatch service will, in near real-time, automatically capture any mentions of brands in the global press and in a broad range of languages.
- **PiiWatch.** PiiWatch is an automated real-time data leak discovery and extraction module. It is designed to search and discover personally identifiable information related to staff, customer and supplier networks.
- **Intellimetrics.** With the power of natural language search terms users can create historical trend maps from billions of data points across every data store available in the platform. They can create real-time content monitors for their dashboard or embed them right into their custom user reports. Intelligent Metrics can be applied to any integrated external data feeds.

To exchange data with any system, the ThreatBridge middleware is used, which allows converting data to whatever format or formats are required by appropriate APIs.

7.3.5. EclecticIQ

EclecticIQ Platform (https://www.eclecticiq.com) is a threat intelligence platform that empowers threat analysts to perform faster, better, and deeper investigations while disseminating intelligence. It connects and interprets intelligence data from open sources, commercial suppliers and industry partnerships. EclecticIQ Platform provides an on-premise solution that gathers data from different sources, configures intelligence feeds, automatically enriches data and builds integration pipelines to IT security controls.

EclecticIQ Platform's workflows include support for a complete range of operational use cases for analysts working within threat intelligence practices at enterprises in high-risk industries. Designed for the real-world activities of CTI analysts, it provides a core set of workflows within a single collaborative workspace. Using these workflows, analysts within SOCs, CERTs, fusion centers, intelligence teams and threat hunting teams can quickly discern actionable and relevant intelligence, collaborate with other analysts, update enterprise security controls and share information with external communities.

The built-in integration capabilities within the EclecticIQ Platform provide enterprises with the flexibility to connect with top providers of threat intelligence and centralized sources of technical data, as well as a full range of IT security solutions deployed within the enterprise. Integration also extends to ISACs and other information-sharing groups using STIX/TAXII standards and other data formats.

EclecticIQ Platform relies on open standards and technologies. Thus, since its inception, EclecticIQ has been an active contributor to the threat intelligence development crowd, notably through their open tools OpenTaxii and Cabby, but also through their contribution to the OASIS standardization body (https://www.oasis-open.org). Additionally, the EclecticIQ Platform provides a robust API and SDK to its community of developers that allows them to own, use, modify, reroute, and eventually transcend it. In more detail, developers can supercharge their EclecticIQ Platform and extend the reach of their intelligence practice by designing new enrichment scenarios, ingesting and exporting of new data sources, and more expressive data model support. The available SDK makes it fast and efficient to develop with EclecticIQ Platform using constructs and syntax familiar to intelligence developers. The SDK is written on top of the EclecticIQ API and provides a complete coverage of the available REST API with documentations, examples, and tools.

7.3.6. iDefense by Qualys

VeriSign and Qualys collaborated to integrate VeriSign iDefense Security Intelligence Services with the QualysGuard Vulnerability Management (https://www.qualys.com/apps/vulnerability-management/) data scanning service. VeriSign iDefense provides access to cyber intelligence related to vulnerabilities, malicious code and geopolitical threats, and thus helping organizations to prevent evolving threats and vulnerabilities. The data for VeriSign iDefense feeds are collected from monitored systems and applications from vendors. Using both human and automated techniques, iDefense filters, analyzes, categorizes and prioritizes the resulting information according to organization relevance, severity and criticality. Thus, it facilitates the procedure of proactively protecting a network, applying customized threat intelligence based on the geographical and contextual needs and provides access to its vulnerabilities knowledge base.

QualysGuard Vulnerability Management is a cloud-based service that provides immediate visibility into where IT systems might be vulnerable to the latest cyber threats and how to prevent them. Additionally, it continuously identifies threats and monitors abrupt changes in networks before they evolve into breaches. QualysGuard Vulnerability Management is built upon a centralized architecture, where all scanning and network mapping data are stored in QualysGuard SOC, while being secured by four layers of protection: SSL encryption, firewalls, IDS sensors and strong encryption methods. Users log into their account, and via an encrypted tunnel, to their own QualysGuard SOC. From the UI provided, users are able to scan both externally or internally their own public or private IP addresses. The external scanning is achieved with a QualysGuard Internet Remote Scanner and through the QualysGuard SOC to the IPs to be scanned. The internal scanning requires an internal appliance which connects to the QualysGuard SOC and receives the actions to follow for the IP addresses to be scanned. Finally, the QualysGuard Vulnerability Management scanning procedure consists of the following phases:

- **Discovery.** QualysGuard learns the nodes of the scanned network, their OS, how they are segmented, etc, and it is also able to provide a visualization of them to the user.
- **Prioritize assets.** User assigns asset groups to the nodes, to better organize them, instead of referring to IP addresses. This assignment could be done depending on the importance value of the nodes or on the business impact for each group.
- **Assessment.** QualysGuard is scanning for vulnerabilities and sends them to QualysGuard SOC for processing. Information discovered during this scan, might not be extremely consumable in the raw scan format.
- **Reporting.** Formatting the data of the raw scan, in a manner that makes sense to an intended audience. The scan results pass through a filter in order to focus on a specific set of information (e.g., a report on how many vulnerabilities have been resolved in the last week). These reports can be altered as necessary, in order to contain information relevant to the audience viewing them.
- **Remediation.** Resolving vulnerabilities found in the assessment phase. In this phase, QualysGuard will point the user to the solution if one exists. A remediation policy might be built, which will create tickets for a specific set of vulnerabilities.
- **Verification.** Verifying that the originally found vulnerabilities, are now fixed and remediated as they were supposed to. In order to check if the remediation was effective, QualysGuard repeats the scanning and verifies that those vulnerabilities are now resolved.

With the integration of VeriSign iDefense and QualysGuard Vulnerability Management [79, 80], iDefense provides the ability to QualysGuard Vulnerability Management to discover zero-day vulnerabilities and create scan signatures for each one of them. Additionally, the correlation of iDefense vulnerabilities and Qualys vulnerability scan data, supports automated vulnerability prioritization based on severity, business impact and relevance to the organization. Furthermore, the integration with the Vulnerability Management solution improves users' ability to respond to emerging threats, making iDefense intelligence more immediately actionable. Moreover, security teams can now prioritize patch deployments and remediation efforts, particularly between vulnerability scan cycles of their networks. Finally, users are provided with a larger knowledge base deriving from both iDefense and QualysGuard Vulnerability Management.

7.3.7. iSight by FireEye

FireEye's iSIGHT (https://www.fireeye.com/solutions/isight-cyber-threat-intelligence-subscriptions.html) TIP is a subscription-based service that delivers actionable intelligence, to customer-owned FireEye and non-FireEye appliances, against new and emerging cyber threats by gathering adversary-focused intelligence across the extended cyber-attack lifecycle. The collected and delivered intelligence gives context to volumes of data regarding global threats, and helps users identify threat actors and indicators of network and system breaches. To do so, FireEye's iSIGHT gathers intelligence from millions of virtual machines that are deployed on networks around the world and stores identified threat data in a centralized data repository that comprises the foundation for all iSight intelligence.

The collected intelligence feeds are diversified [34], based on different security roles assumed by the user, to:

- **Tactical intelligence.** Aimed at tactical/technical users and involves a rich data feed and alerting but no intelligence reports/digests.
- **Operational intelligence.** Aimed at SOC personnel and IR teams and involves actionable context (e.g., threat actor, malware profiles), data feeds and alerts, and a threat prioritization module.
- **Fusion/executive intelligence.** Analytic reporting aimed mainly at SOC and IR personnel that actively search for adversaries, as well as condensed, non-technical reports aimed at CISOs and executives.
- **Vulnerability intelligence.** Aimed at IT personnel and involves information on patches and active/emerging threats.

The offered service supports open standards (e.g., intelligence reports are STIX-formatted), established practices (e.g., responses default to JSON format, but XML formatting may also be set and used), full-fledged search capabilities (e.g., advanced and pivot search around IoC), up-to-date periodical vulnerability information (e.g., by programmatic access to vulnerability data, and associated attributes, including CPE, CVE ID, CVSS scores), and easy integration (e.g., SDK/API with code samples).

The iSight Critical Infrastructure [26] is a specialized TIP used for protecting critical infrastructures, industrial control systems, industrial IoT, and cyber physical systems by providing (subscription-based) insight into the intent and capabilities of threat actors, and cross-referencing identified threats against events from the existing information and operational technology. Its integration with FireEye's Threat Analytics Platform [26], that monitors operational network traffic, and the utilization of specialized industrial security firewalls provide a minimally-invasive protection solution that is mainly built upon intelligence.

7.3.8. MANTIS by Siemens

The *Model-based Analysis of Threat Intelligence Sources* (MANTIS) framework (http://django-mantis.readthedocs.io) was introduced by Siemens as a basis for managing CTI expressed in a number of standards including STIX, CybOX, OpenIoC, IODEF, and others. MANTIS consists of a number of Django (Python-based web application framework) apps [44] and provides an information repository into which cyber threat intelligence received in a variety of standards/formats can be imported and used for browsing, filtering, searching and correlation. The goal of the MANTIS project is to stage and efficiently correlate threat intelligence feeds as attributed graphs of observables. To this end, it utilizes a semi-automated tool that leverages observable linkage and allows users to compare relationships between different observables, rather than explicit query of known observables (signature). This feature makes MANTIS especially useful for threat intelligence with sparse data profiles.

MANTIS is currently used as (a) an information base for maintaining threat intelligence generated or received by companies and organizations and (b) as a correlation and exchange mechanism of threat information by resorting on some of the currently available standards. Due to the adaptability and configurability of the data importer that comes with the MANTIS Framework other structured data (sources) may be easily integrated into the framework. The centralized architecture and the open source nature of the project [28] offer a number of advantages including (a) lowering the entrance barrier for organizations and teams (esp. CERT teams) in using emerging standards for CTI management and exchange, (b) providing an example implementation that may function as a basis for research and community-driven development of cyber-threat intelligence management, (c) aiding discussions on emerging (and older) standards regarding expressivity, functionality, and tool development.

Currently MANTIS offers a suite of tools and functionalities that include login and menu construction, importing threat intelligence expressed in a number of formats and standards, viewing and filtering threat information objects (either in the original or in a JSON representation) along with their revisions, searching for threat information and result viewing, user data and configuration editing, and an admin interface. Shortcomings of the framework include the sluggish maintenance and updating, the adoption of preterite versions of standards (e.g., only the XML-based STIX (1.0) is currently implemented), and the need for conversion to a Python project (as opposed to the current Django application).

7.3.9. Emerging Threat Intelligence by Proofpoint

Proofpoint's Emerging Threat Intelligence (https://www.proofpoint.com/us/products/et-intelligence) combines actionable up-to-the-minute IP and domain reputation feeds with a database of globally observed threats and malware analysis, in order to give the security professional the necessary intelligence to proactively stop malicious attacks and provide the context needed to investigate them.

Emerging Threat Intelligence leverages one of the world's largest active malware exchanges, victim emulation at massive scale, original detection technology and a global sensor network.

Emerging Threat Intelligence provides threat intelligence feeds to identify IPs and domains involved in suspicious and malicious activity. Separate lists are created for IP addresses and domains; each IP and domain can be classified into over 40 different categories and assigned a confidence score for each category. The scores indicate recent activity levels and are aggressively aged to reflect current conditions. Multiple formats are supported, including TXT, CSV, JSON and compressed.

The threat intelligence feeds can be directly fed to existing SIEM tools such as Splunk, QRadar and ArcSight and various TIPs. Finally, Emerging Threat Intelligence is directly available for use through the ThreatStream by Anomali.

7.3.10. Recorded Future

Recorded Future is a SaaS product (https://www.recordedfuture.com/). The technology behind Recorded Future is the Threat Intelligence Machine, which utilizes machine learning and multilingual NLP, in order to continuously analyze cyber threat data from a variety of sources (such as social media, paste sites, hacking forums, marketplaces) on the surface and the dark web. Specifically, this technology automatically and continuously harvests information from this breadth of sources in parallel, correlating and combining them in order to provide contextualized intelligence in real time, and highlight IoCs about vulnerable hosts and technology. The machine processes the incoming information 24/7 and combines it with the Recorded Future knowledge base. This intelligence can uncover specific risks relevant to each organization. In addition, following the machine learning approach for over more than 20 billion data points, the natural language processing can extract hundreds of facts per second and it has learned to drastically reduce the noise of potential false positives. Furthermore, this machine learning technology achieves a 99% precision in predicting future malicious IP addresses and generates predictive models that identify the likelihood of product vulnerabilities being exploited.

Utilizing this technology, Recorded Future provides a vast collection of real-time threat intelligence, since it continuously processes billions of data points in multiple languages, from technical, open and closed sources (e.g., dark web). In addition to this real time monitoring, Recorded Future supports immediate custom alerts, relevant to the user's cyber threat surface, brand and infrastructure. Moreover, users can centralize threat data of their concern from the Recorded Future knowledge base, view relevant insights according to their corporate profile, and get personalized solutions depending on the industry they belong to. Currently, Recorded Future supports specific solutions for financial, healthcare, retail, energy and government industries. Several other key features of Recorded Future include:

- **Intelligence cards.** Recorded Future Intelligence Cards contain structured real-time cyber threat information such as malicious or targeted IP addresses, vulnerabilities, and malware, in order to drive faster analysis and confident decisions.
- **Threat views.** Taking advantage of Recorded Future Threat Views capabilities, users can customize the visualizations of the threat landscape to view emerging attackers, methods and indicators. By fine-tuning the Threat Views, users are presented with trending intelligence relevant to their brand, technologies and industry.
- **Search.** Recorded Future provides selective search capabilities through customizable querying that lets users quickly refine results to get the most relevant information.

Finally, through the API, Recorded Future can be integrated with existing security technologies, delivering context for indicators of compromise, malware, vulnerabilities and an accompanying risk score, and also support data formatting in standards like STIX/TAXII.

7.3.11. Soltra Edge by NC4

Soltra Edge (https://www.soltra.com/en/products/soltra-edge/) is a platform for sharing and automating CTI within an organization and the outside world. Soltra Edge alters the current paradigm by (a) accelerating the process of risk detection to action taking, (b) allowing users to make decisions and mitigate threats more quickly, and (c) increasing an organization’s operational capacity to manage threats. Soltra Edge is able to act as a central cyber threat intelligence repository, by aggregating data from internal and external sources, normalizing them in STIX format and managing to create actionable alerts. This central intelligence repository filters and controls threat information that is sent to other applications and devices in an organization's cybersecurity stack and avoids unnecessary tasks and alerting by incorporating new threat identification and defense strategies in the form of Soltra Edge filters and controls. In summary, Soltra Edge can be used [70] to:

- Share threat intelligence (send and receive) with ISACs, ISAOs, industry associations, communities, trust groups, DHS and others.
- Serve as a router of threat intelligence to existing security applications and devices such as SIEMs or firewalls.
- Connect and interoperate with software from other vendors such as IBM, LogRythm, AlienVault, HPE, Splunk, Intel Security, Cisco, Palo Alto, Phantom, Tanium, Tripwire, Carbon Black, ServiceNow, ThreatQuotient (see Section 7.3.13. ThreatQ by ThreatQuotient), Anomali (see Section 7.3.14. ThreatStream by Anomali).
- Provide robust search and tagging of cyber threat data.

Soltra Edge manages CTI by using (all eight core) STIX constructs and incorporates of the TAXII, CybOX, and OASIS threat intelligence sharing standards to (a) ensure easy interoperability with other applications and devices that are compliant with those standard, (b) allow for integration with non-standard data sources, and (c) facilitate the sharing of CTI in both intra- and inter-organization setups. Soltra Edge currently offers both an on-premises version that may be installed on virtual appliances or a physical server and a SaaS version that runs in NC4's secure data center. Moreover, Soltra Edge features a rich set of available tools for managing CTI data such as a data summarization dashboard, a full-fledged search engine for different object types stored in the database, human viewing and authoring tools (with automated correction capabilities) for STIX documents, manual and automated threat intelligence data publication, trust group creation for data sharing and information control, various plug-ins and an effective user management and access control mechanism. Finally, although Soltra Edge was primarily designed for general-purpose threat intelligence collection and sharing, it has been also tested and successfully deployed in the financial sector. To this end, it serves as the central repository for the only industry forum for collaboration on critical security threats in the global financial services sector, FS-ISAC (Financial Services Information Sharing and Analysis Center).

7.3.12. ThreatConnect

The ThreatConnect Platform (https://www.threatconnect.com) offers an integrated set of different threat intelligence tools that may be utilized (independently or in combination) by companies depending on their needs and size. The tools offered include:

- threat intelligence gathering (ThreatConnect Identify),
- threat data management, aggregation, and orchestration (ThreatConnect Manage),
- threat data analysis and prioritisation (ThreatConnect Analyze), and
- complete suite (ThreatConnect Complete).

ThreatConnect Identify provides actionable threat intelligence by monitoring more than 100 open source feeds, crowdsourced intelligence from different security communities (including the company's security and research teams), and optionally partner companies involved in cyber intelligence [75]. Threat intelligence sharing, collaboration with other stakeholders, and actionable integrations with SIEM software, firewalls, and other endpoint protection options are facilitated by resorting to STIX/TAXII standards. The platform offers a fully integrated TAXII client and all platform users operate over STIX-formatted threat intelligence either with their own cloud or with an on-premises ThreatConnect instance. ThreatConnect

Manage builds upon the collected threat intelligence and allows users to automate (part or all of) the threat data management processes; this includes actions like threat data augmentation (either manually or semi-automatically with information like adversary activity, country of origin, phase of intrusion), indicator sharing, the application of remediation or mitigation strategies from defensive tools, or task monitoring. Finally, ThreatConnect Analyze provides a central place to monitor team tasks, analyze (threat) data, and use the security arsenal at hand.

Although the ThreatConnect Platform is primarily meant for general-purpose threat intelligence collection and sharing, it may be customized to fit the needs of vertical sectors. Lately, special emphasis has been put in the utilisation of ThreatConnect in medical and health verticals, where the risks associated with compromise are often significantly augmented as patient care and personal information are at stake. Generally, medical and health organizations (e.g., in the pharmaceutical sector), face a variety of threats that are inherent to the services they provide and the data they safeguard (e.g., ransomware, intellectual property theft, or intelligence collection to enable domestic drug production). To this end, the ThreatConnect Platform collects and identifies notable threats, incidents, campaigns, tags, and communities that are pertinent to medical and health organizations within the ThreatConnect ecosystem and leverages them for use in the medical or health sectors.

7.3.13. ThreatQ by ThreatQuotient

ThreatQ (https://www.threatq.com/threat-intelligence-platform/) is a CTI platform designed to accelerate security operations by leveraging an integrated self-tuning threat library to automatically score and prioritize threat intelligence based on user-set parameters. This library constitutes a central repository within the actual TIP and combines structured and unstructured threat data to provide relevant and contextual threat intelligence. Over time, the library self-tunes, enabling situational understanding, better decision making and automated actions that accelerate security operations. Threat prioritization is calculated across many separate external and internal sources to deliver a threat score that is also based on the aggregated context. The use of the threat library is a key design choice that aids at removing noise, reducing the risk of false positives and enabling users to focus on the actual threats.

The ThreatQ platform itself constitutes an open, extensible and robust ecosystem of tools and modules [78] that (a) provides STIX/TAXII and OpenIoC support for the management (e.g., importing, aggregation) of over 200 different threat data sources, feeds and integrations, (b) can easily integrate with existing enrichment and analysis tools and workflows developed in-house or by third-parties -such as the open source IDS/IPS/NSM engines Snort (https://www.snort.org) and Suricata (https://suricata-ids.org), and (c) is highly customizable through an available SDK. The ThreatQ platform is available in four different modes to fit on the network design and user needs: (a) on-premises, providing maximum security and complete control over the network data, (b) cloud-based, offered as a service, (c) virtual instance, available in software only OVA distributions for virtual machine deployment, and (d) dedicated appliance, for meeting increased performance requirements. Noteable use cases for the ThreatQ platform [78] are:

- **Threat data aggregation and curated intelligence.** ThreatQ combines, normalizes and contextualizes threat data from external/internal sources and leverages/prioritizes them to through context.
- **Identify and investigate attacks/adversaries.** Users can utilize campaign, malware and indicator knowledge to identify related attacks and adversaries, investigate and track spear phishing attacks,

or support scoping and breach investigation by correlating artifacts of an investigation with the system's threat library.

- **Strengthen the defensive arsenal.** Leverage existing firewall, IDS, IPS, or SIEM with accurate and relevant threat data, assist threat hunting by proactively searching for (yet) unobserved malicious activity.

7.3.14. ThreatStream by Anomali

ThreatStream (https://www.anomali.com/platform/threatstream) collects threat intelligence data from hundreds of sources that include:

- STIX/TAXII feeds.
- Open source threat feeds.
- Commercial threat intelligence providers.
- Structured and unstructured intelligence.
- ISAC and ISAO shared threat intelligence.

Following, ThreatStream (a) normalizes feeds into a common taxonomy, (b) removes duplicates across data feeds, (c) removes false positives via machine learning algorithms, (d) enriches data with Actor, Campaign, etc., (e) adds context from WHOIS, PassiveDNS, and other lookup information and (f) associates related threat indicators. ThreatStream provides tools to help analysts and SOC teams and respond to threats. Particularly, it includes features such as:

- Phishing that extracts indicators from suspected emails.
- Sandbox that detonates malware and extract relevant indicators.
- Brand monitoring that detects of brand abuse.
- Threat investigation engine with analyst workflows.
- Threat bulletin creation, management, and collaboration.

Additionally, trusted circles within the ThreatStream platform ensure that users can participate seamlessly in two-way sharing. Company-proprietary information can be kept private to guarantee confidentiality of shared information.

7.4. Complete threat intelligence solutions

7.4.1. DeepSight Intelligence by Symantec

Symantec DeepSight Intelligence[29] is a SaaS product which provides access to the cloud-hosted CTI platform of Symantec; this platform encompasses both DeepSight Adversary Intelligence and DeepSight Technical Intelligence [68]. DeepSight Adversary Intelligence helps users understand their threat environment, including cyber espionage, cyber-crime and hacktivist threats, via the monitoring of various sources. Specifically, users get a better understanding of all information about the actors and groups behind several attacks, their motivations, exploited vulnerabilities and malware utilized.

DeepSight Technical Intelligence provides all information about cyber threats stored or recently discovered by DeepSight:

[29] https://www.symantec.com/services/cyber-security-services/deepsight-intelligence

- **Vulnerability intelligence.** DeepSight provides inclusive vulnerabilities knowledge across many technologies from various vendors. In addition to that, DeepSight also provides risk scores, impacted products, patch availability and exploits along with each vulnerability indicator.
- **Network reputation.** DeepSight houses information about ownership, reputation, and event data on IPs, domains or URLs that have been observed to act maliciously in the past. Thus, users are able to disable inbound and outbound communications with these malicious IPs, domains or URLs in order to ensure the security of their networks.
- **Security risk/malcode.** DeepSight analyzes viruses, worms, trojans, adware, spyware and other potentially harmful files and applications, to provide behavioral characteristics of these threats and help users determine how to take corrective actions.
- **File reputation.** DeepSight provides file reputation intelligence for billions of files known to its global information network, to identify, analyze and stop the distribution of a malware entering the user's network through a known malicious file.

Combining the two aforementioned intelligence forms, DeepSight portal [69] provides an environment of cyber threat alerting, that can be easily tailored according to the needs of each user. Through the vulnerability management option, users can create alerting profiles depending on the technology they use (by providing vendors, products and versions). Then, they can define the method of the alert delivery (via email or SMS). Lastly, users choose the cyber threat type (vulnerability, malicious code, security risk, and more) that they want to monitor and then define the critical threshold settings, such as urgency, severity, impact, etc. After following the described procedure, every time a vulnerability affects the profiled technology, users will receive detailed reports about the threats, along with recommendations on how to mitigate the risk.

Another key feature of the DeepSight portal, is the provision of an environment about cyber threat insights. Through this environment, users can view statistics about past cyber threat incidents and also lookup in the DeepSight knowledge base for specific malicious IPs, domains or URLs, suspicious files and more.

Finally, DeepSight Intelligence offers two more options for delivering their cyber threat intelligence, other than the DeepSight portal; data feeds and APIs for intelligence automation [68]. Formatting their data into XML for the data feeds and into JSON through the API, helps DeepSight to integrate with other existing security solutions.

7.4.2. Falcon Intelligence by CrowdStrike

CrowdStrike's Falcon Intelligence (https://www.crowdstrike.com/products/falcon-intelligence/) offers an in-depth and historical understanding of adversaries, their campaigns, and their motivations. Falcon Intelligence reports provide real-time adversary analysis for effective defense and cybersecurity operations. More specifically, Falcon Intelligence delivers:

- Immediate alerting and warning of new adversary activity.
- Weekly, periodic and quarterly strategic, operational and technical reports.
- Tailored intelligence which provides proactive alerting on keywords and expressions.
- APIs, feeds, and rules for easy integration with existing infrastructure (SIEMs, TIPs, and more).

Also, by combining comprehensive analysis with threat indicators, Falcon Intelligence can (a) provide visibility into future threats, (b) help understand adversary capabilities, motives and tradecraft, (c) optimize the organizations resources to determine targeted intrusions versus. broad-based cybercrime attacks, saving time and focusing response efforts on critical threats and (d) implement effective countermeasures against emerging threats with timely, thorough reports and API feeds.

The Falcon Intelligence API allows customers to automate the consumption of indicator data collected by the CrowdStrike intelligence team and obtain attribution information for any supported indicator type. Subscribers can integrate cyber threat intelligence and information into their existing security architecture to quickly gain insight into advanced malware and targeted attacks by leveraging the offered API. The web-based API enables collection and querying of hashes, domains, IP addresses, and much more in JSON. Consequently, conversion into standard, proprietary, and device-specific formats can be easily accomplished with user scripts to accommodate integration with a wide variety of devices.

7.4.3. InTELL by Fox-IT

Fox-IT (https://www.fox-it.com) is one of Europe's largest specialized cyber-security companies and one of the first to establish a SOC for cyber-security in Europe, with an active involvement in many high-profile IR cases. The InTELL platform (https://www.fox-it.com/intell) developed by Fox-IT is developed to serve as the gateway to all underlying systems in the cyber threat management ecosystem of Fox-IT [11]. The platform abstracts the technical and operational information for tactical and strategic management; the user and access control mechanism determines the functionality, information form and abstraction that is presented to each user type. The InTELL platform offers detailed threat analytics with integrated workflows, threat research facilities for easier investigation of security incidents, and platform adaptation according to the identified threats. Moreover, it offers online collaboration tools that allow on-premises analysts to collaborate with peers within Fox-IT or other partners.

The InTELL platform tracks global criminal activity and sorts the collected threat intelligence based on actor attribution and context. Apart from harvesting typical threat information, InTELL provides a global picture of trends, geographical activity, and actors (along with identified links to campaigns, tactics, procedures and individual IoCs). InTELL provides actionable threat intelligence by utilising a contextual STIX feed containing IoCs, TTPs, campaign and actor attribution. The feed is fully compatible with other third-party STIX-capable solutions, such as Soltra Edge (see Section 7.3.11. Soltra Edge by NC4) and provides threat information that can be digested by SIEMs, threat platforms, or risk engines and may be utilized to automate an organization's SOC or its detection processes. Overall, the platform's open and modular architecture allows its integration with existing information sources and its expandability with a rich set of options including (a) the Managed Intelligence Service module [11] that provides proactive cyber vigilance for vital organizational assets (e.g., intellectual property or reputation), and (b) customizable network, endpoint or log detection modules.

Finally, although Fox-IT has a long-term experience (and customized products such as DataDiode) in securing classified government systems, InTELL's threat intelligence solutions are not targeted towards critical and industrial infrastructures and vertical sectors.

7.4.4. Talos Threat Intelligence by Cisco

The Talos threat intelligence system (https://www.talosintelligence.com/) by Cisco provides actionable threat response management by consuming threat intelligence from multiple sources, combining the collected information with local contextual information, proposing and accepting COAs, determining which response systems may carry these actions, and subsequently putting them into effect [2]. The system is implemented on top of the Cisco Platform Exchange Grid [9] and is currently under testing for inclusion in Cisco's cybersecurity umbrella. The threat intelligence core of the system adopts a pub/sub model for matching COAs with proposed actions and resorts to a decentralized architecture aiming at robustness, scalability, and efficiency. To do so, it leverages the pub/sub messaging protocols, federation and security capabilities of the Extensible Messaging and Presence Protocol (XMPP), widely used in collaboration systems. XMPP provides the necessary scalability component and constitutes an open source standard able to extend its language to accommodate different data models and protocols -including those within the security context, such as IODEF (see 5.6. IODEF (Incident Object Description Exchange Format)). The decentralized and loosely-coupled architecture of the core mechanism allows each response system to determine independently whether it can carry out a specific COA or not. To this end, each response system subscribes to the COA feed, while the threat intelligence system publishes the desired COAs. Subsequently, under this model, each subscriber independently determines its capability of carrying out the published action. Redundant responses (typically categorized as investigation, mitigation, and remediation) are handled by assigning the COA to the first responder(s) that can carry out the proposed COAs. Apart from the (asynchronous) pub/sub functionality, the system offers also (synchronous) direct querying services.

STIX is used as the language of choice to express and share not only threat information such as IoCs and incidents, but also actionable details such as suggested, requested and taken COAs. The response actions are divided into three main categories: (a) investigate - obtain more information about a threat, (b) mitigate - block but do not eliminate a threat, and (c) remediate - fix or eliminate a threat. The threat intelligence system may invoke any (or a combination) of these actions from the response systems capable of delivering the functionality as identified by the pub/sub model.

Finally, special emphasis has been given on supporting strong access control on threat intelligence to avoid inappropriately sharing sensitive data. For example, the threat intelligence system needs to be authorized to consume and process STIX information from threat intelligence providers.

Overall, the scalability of the threat intelligence system is directly derived from the architectural choices that involve a loosely-coupled decentralized architecture bundled with a pub/sub model for component interactions and information dissemination. Currently, in a single out-of-the-self server with a few hundred subscribers, the system can process more than 1.5K direct queries and 200 events per second [2]; higher rates are possible in a clustered deployment, by having servers route messages among them according to the pub/sub protocol.

7.5. Summary

This section provides a short summary of the platforms presented in the previous sections, following their classification (shown in the first column of the table below) based on the supported functionalities.

Table 30. Market summary

Section	Market solution	Vert. sector support	Architecture	Pub/sub	Standards used	SaaS / on-prem
7.1.1	BreachAlert	No	Centralized	Yes	N/A	N/A
7.1.2	DyTA	Yes	Centralized	No	N/A	SaaS

Section	Market solution	Vert. sector support	Architecture	Pub/sub	Standards used	SaaS / on-prem
7.1.3	F5 labs	No	Centralized	N/A	N/A	SaaS
7.1.4	Flashpoint Platform	Yes	Centralized	Yes	N/A	SaaS
7.1.5	Intel 471	No	Centralized	Yes	N/A	SaaS
7.1.6	LookingGlass CTI	No	Centralized	Yes	N/A	SaaS
7.1.7	Norse Intelligence	Yes	Decentralized	Yes	N/A	SaaS, on-prem
7.1.8	SearchLight	No	Centralized	Yes	STIX/TAXII	SaaS
7.1.9	Security Ratings	No	Centralized	Yes	N/A	SaaS, on-prem
7.1.10	Wapack Labs	No	Centralized	Yes	N/A	SaaS
7.1.11	ZeroFOX	Yes	Centralized	No	N/A	N/A
7.2.1	CRITs	No	Centralized	Yes	STIX/TAXII	SaaS
7.2.2	Enclaves	No	Centralized	Yes	STIX/TAXII	SaaS
7.3.1	ActiveTrust	No	Centralized	Yes	STIX/TAXII	SaaS, on-prem
7.3.2	BrightCloud Services	No	Centralized	N/A	Proprietary API	hosted, on-prem
7.3.3	Cyber Advisor	Yes	Centralized	Yes	STIX/TAXII	SaaS
7.3.4	Cyjax Tech. Stack	Yes	Centralized	Yes	Conversion into standards via API	SaaS
7.3.5	EclecticIQ	No	Centralized	Yes	STIX/TAXII	on-prem
7.3.6	iDefense	No	Centralized	Yes	Proprietary API	SaaS
7.3.7	iSight	Yes	Centralized	Yes	STIX/TAXII, other standards via APIs	SaaS
7.3.8	MANTIS	No	Centralized	Yes	STIX/TAXII, CybOX, OpenIoC, IODEF	on-prem

Section	Market solution	Vert. sector support	Architecture	Pub/sub	Standards used	SaaS / on-prem
7.3.9	Proofpoint ETI	No	Centralized	Yes	STIX/TAXII	SaaS
7.3.10	Recorded Future	Yes	Centralized	Yes	STIX/TAXII	SaaS
7.3.11	Soltra Edge	Yes	Centralized	No	STIX/TAXII, CybOX	SaaS, on-prem
7.3.12	ThreatConnect	Yes	Centralized	No	STIX/TAXII	SaaS, on-prem
7.3.13	ThreatQ	No	Centralized	No	STIX/TAXII, OpenIoC	SaaS, on-prem
7.3.14	ThreatStream	No	Centralized / Cloud	N/A	STIX/TAXII	SaaS, on-prem
7.4.1	DeepSight Intelligence	No	Centralized	Yes	Conversion into standards via API	SaaS
7.4.2	Falcon Intelligence	No	Centralized	Yes	Conversion into standards via API	SaaS
7.4.3	InTELL	No	Centralized	Yes	STIX/TAXII	SaaS
7.4.4	Talos Intelligence	No	Decentralized	Yes	STIX/TAXII, CVRF, OVAL, others	SaaS

In what follows, we outline some key findings emanating from the market analysis conducted in the previous sections and also from the summary table above.

- **STIX is the de-facto standard for describing threat intelligence**. The landscape of standards available to describe threat intelligence is rather small compared to the number of available systems. Our market analysis showed that most CTI sharing solutions rely on STIX/TAXII, and a few of them support also other standards like OpenIoC, IODEF, and CybOX. About half of the examined solutions provide direct import and export capabilities supporting STIX/TAXII, while only a small fraction exposes a proprietary API (that in many cases also accepts standardized CTI formats and languages). In detail, sixteen market solutions rely on STIX, two on OpenIoC, two on CybOX, one on IODEF, and one on CVRF and OVAL. Moreover, a number of market solutions support multiple standards either directly or through a proprietary API. Overall, STIX is the most commonly used approach and can be considered as the de-facto standard for describing threat intelligence. It builds upon the CybOX, CAPEC, MAEC and CVRF standards, and provides a unifying architecture tying together a diverse set of cyber threat information [82]; several market solutions report to take into account most or all eight core cyber threat concepts.

- **Vertical sector support is not very common.** Although the CTI landscape is rather large, less than one third of the examined solutions take into account vertical sectors; different market solutions focus on different vertical sectors without a clear stand out. Thus, the offered CTI solutions for verticals range from CIIs to health/pharmaceutical and financial domains. It is also worth noting that even companies with background and expertise in providing security solutions for vertical sectors, do not offer CTI products in the corresponding domains.
- **Centralization dominates the market**. Our market analysis has shown that despite the advertised volume of monitored CTI sources and the number of handled incidents and generated reports, centralization is the dominant architectural solution for the vast majority of the examined systems. Only a small fraction (around 10%) of the examined market solutions is designed with decentralization in mind, while all offered solutions make strong scalability claims. It is also worth noting that while more than two-thirds of the systems offer CTI as a service, one a handful of them has adopted a decentralized architecture.
- **Pub/sub is becoming a standard, but content-level dissemination is still in its infancy.** More than two-thirds of the existing solution offer some type of notification service for the involved stakeholders; this service is typically a channel-based or type-based alerting module that is responsible for disseminating important CTI-related information to all parties involved. However, content-based alerts are adopted by very few market solutions, probably due to performance reasons (i.e., matchmaking in content-based information dissemination is significantly more expensive than channel/type-based). It is also worth noting that pub/sub is mostly prevalent to data feed providers and complete CTI solutions that have to either handle data streams or to provide a full set of integrated threat intelligence services.
- **CTI as a service.** With the advent of the cloud, an increasing number of companies rely on services (or hypervisors and related technologies) to cover for their computing, storage, and (lately) security needs. This trend has given rise to concepts like security-as-a-service; an outsourced SaaS wherein an outside company handles and manages another company's security without the need to deliver any security solution locally. This new business model is typically subscription-based and is currently supported by all major security players and a growing number of smaller (and more focused/specialized) CTI-related companies. Our market analysis has shown that two-thirds of the examined CTI solutions are solely SaaS-based, around 25% offer both a SaaS and a local solution, whereas only two companies offer a solely on-premises CTI infrastructure.
- **Sharing is mainly for IoCs.** The majority of the observed systems primarily focuses on the sharing of IoCs to enable the identification of potentially malicious activities. While the OpenIoC standard is primarily designed to share them, only a handful of solutions uses it; most rely on STIX's Observable and Indicator constructs to describe IoCs (*see* the first two rows of Table 31 at the next page).
- **Collection and storage of CTI are the most prominent actions supported.** The majority of the examined solutions primarily focus on data collection and often neglect other activities of the intelligence lifecycle. To this end, most CTI solutions are closer to data warehouses than intelligence sharing. This becomes more apparent by the fact that only a small number of platforms provide interfaces for third party tools to facilitate sharing or further analysis. Finally, all CTI solutions provide tools that target human analysts (i.e., browsing, filtering, searching, and visualization) and also basic or advanced (automated) analysis of CTI information.

8. Recommendations

Although several standards exist for the sharing of cyber-threat related information, STIX is clearly shown to be industry's preferred exchange mechanism as key findings of 7. Current market situation highlight. This is also aligned with the outcomes in the reports of ENISA, where CERT, CSIRT and LEA communities also find STIX to be a suitable candidate for a sharing mechanism [18, 20]. Therefore, STIX enjoys the acceptance from all the project's stakeholders. This is due to the fact that STIX allows for the accurate description of information, including the source of data (which is necessary for reasoning about trusting the data). STIX also allows different parts in the model to be implemented in a step-by-step approach by gradually incorporating part of grown in complexity as shown in the Table 31, where the sharing of simple CTI indicators is performed in most cases.

Table 31. Full spectrum of CTI and STIX

CTI indicators	Full CTI spectrum	What activity are we seeing?
		What threats should I look for on my networks and systems and why?
		Where has this threat been seen?
		What does it do?
		What vulnerabilities does this threat exploit?
		Why does it do this?
		Who is responsible for this threat?
		What can I do about it?

In addition to the above, whenever the need to obtain more detailed information, within Cyber-Trust, in certain contexts, other standards may be embedded in STIX. This is very useful for malware cases or to identify CTI belonging to distributed denial of service attacks, which are amongst the attacks that the Cyber-Trust project is expected to focus on. As shown in 5.4.4. MAEC and STIX, MAEC is such an example where the embedding of native MAEC data into STIX allows capturing detailed structured information alongside broader cyber-threat related information.

By recommending STIX as the sharing mechanism, the next thing to decide is the choice of the sharing platform, where the need to ensure that the chosen sharing mechanism can be supported by *open source products* has driven the work in 6. CTI platforms and tools. The analysis presented in 6.7. General and technical attributes of platforms showed that all sharing platforms support STIX, still, MISP and GOSINT stand out. These two platforms both cover to a good extent the properties imposed by Requirements 3 and 4, but overall MISP has considerably more advantages (see next Table 32) and, thus, it recommended as the sharing platform of the Cyber-Trust project.

Based on the analysis presented in 6.7. General and technical attributes of platforms,6.7. General and technical attributes of platforms although all tools support STIX, the choice of MISP as the sharing platform to use in Cyber-Trust project has considerable advantages (see next table) compared to other alternatives (and primarily GOSINT that also covers to a good extent the properties imposed by Requirements 3 and 4).

Table 32. Strengths and weaknesses of the CTI sharing platforms

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Strengths	● Able to use the IoCs and information to detect and prevent attacks or threats against ICT infrastructures ● Incident handling ● Available threat models ● Creation of your own models ● Sensitivity designation ● Very active support by MISP community	● Collect and standardize structured and unstructured CTI ● Enriched alert data ● Indicator quality, through indicator judging ● No limit to the number of indicator sources	● Comprehensive threat-scoring framework ● Efficiently automating those decisions.	● Has a powerful repository of threats. ● Many tools to support the manipulation of stored data. ● APIs automation	● Big repository containing by data and metadata of simple or combined intrusions ● APIs automation	● Automation by APIs

Platforms	MISP	GOSINT	OPENTPX	Yeti	OPENTAXII	CIF
Weaknesses	● N/A	● Not provide up to date versions of the software ● Package managers may name packages differently	● Do not provide tools for creation of incident attack. ● It is not human readable	● Do not provide tools for creation of incident attack.	● Do not provide tools for creation of incident attack.	● Only observed threats (such as IPs)

A number of key strengths and weaknesses for each platform are illustrated in the Table 32. Apart from meeting the requirements stated in 3. Description of methodology, MISP also allows to designate the sensitivity or classification level of information by means of protocols like the *traffic light protocol* (TLP), something that is relevant to the project as privacy issues need also be taken into account (*see* Annex B). In addition, MISP enjoys a very strong and active community of supporters, from the private sector, public bodies, and other organizations, that keeps extending and improving its functionality (e.g., by creating a series of data models). This list includes a continuously growing number of CERTs, CSIRTs, etc., such as CIRCL (computer incident response center Luxembourg)[30], CERT-BW, GOVCERT.LU, NorCERT, MIL.be, defCERTNL, NCI (NATO Communications and Information) Agency. Due to the above reasons, MISP platform not only meets the requirements set forth but shows a good promise to allow Cyber-Trust platform to contribute to the MISP community by sharing CTI data.

[30] https://www.circl.lu/

9. Conclusions

This deliverable overviewed and critically evaluated existing industry-wide vulnerability reporting and sharing sources, standards, frameworks and platforms to provide recommendations on the approach to be followed in the Cyber-Trust platform.

Initially, sources of data for threat information sharing systems were presented and categorized into internal, community, and external with the purpose of compiling a cataloging inventory that contains elements useful for the purposes of the project. Such elements may include the type of exposed data (e.g., structured machine-readable or unstructured) and query languages, protocols, or services available for data retrieval.

Subsequently, the appropriateness of different vulnerability reporting frameworks for disseminating the identified cyber-threats across different organizations and promoting awareness about emerging cyber-threats were considered. Moreover, issues pertaining to the basic structure, the key elements (i.e., expressiveness, flexibility, extensibility, automation, structuring), and prominent strengths/weaknesses of the presented frameworks were discussed and critically evaluated within the scope of the Cyber-Trust project. Frameworks and languages for supporting expressive content-based subscriptions in the context of specialized pub/sub services for cyber-threat information push were also considered.

The presented frameworks and languages were realized in platform and tool implementations to provide the necessary functionality and enhance standard adoption. The mechanisms for handling structured cyber-threat information for a wide variety of use cases (including those outlined in the project) were also presented alongside important components that include the key characteristics of each platform, the supported observables and schemas, and the adopted standards.

Following, prominent market solutions related to the discovery and management of cyber threat intelligence were reviewed and categorized into services, data feeds, platforms, and complete systems. The main features and characteristics with respect to several different facets including architecture, offered services, standards' adoption, and mode of operation were critically compared for each category to highlight salient market practices that relate to the goals of the Cyber-Trust project.

Finally, based on our analysis, we present our recommendations for the Cyber-Trust project.

10. References

[1] Abuse.ch. (2017) Because malware sucks. [Online]. Available at: https://abuse.ch/

[2] S. Appala, N. Cam-Winget, D.A. McGrew, J. Verma. (2015) An Actionable Threat Intelligence system using a Publish-Subscribe communications model. WISCS@CCS

[3] Abuse.ch. (2018) SSL Blacklist. [Online]. Available at: https://sslbl.abuse.ch/blacklist/

[4] Abuse.ch. (2018) URLHaus Database API. [Online]. Available at: https://urlhaus.abuse.ch/api/

[5] Anomali. (2018) Limo - Free Intel Feed. [Online]. Available at: https://www.anomali.com/platform/limo

[6] Anomali. (2018) Enterprise Threat Intelligence Solutions. [Online]. Available at: https://www.anomali.com/

[7] Botvrij.eu. (2018) Freely available network IoCs. [Online]. Available at: http://www.botvrij.eu/

[8] Blocklist.de. (2018) fail2ban reporting service. [Online]. Available at: https://www.blocklist.de

[9] Cisco Platform Exchange Grid. (2017) [Online]. Available at: http://www.cisco.com/c/dam/en/us/products/collateral/security/identity-services-engine/at_a_glance_c45-728420.pdf

[10] CERT Tools. (2018) IntelMQ is a solution for IT security teams for collecting and processing security feeds using a message queuing protocol. [Online]. Available at: https://github.com/certtools/intelmq

[11] Fox-IT (2016) Cyber Threat Management platform: defense against known and unknown threats. White paper. [Online]. Available at: https://www.fox-it.com/en/wp-content/uploads/sites/11/Fox-IT_CTMp_brochure_February_2016.pdf

[12] M. Schiffman. (2011) “The common vulnerability reporting framework,” *Cisco Systems, Inc., Tech. Rep.* [Online]. Available at: http://www.icasi.org/wp-content/uploads/2015/06/cvrf-whitepaper.pdf

[13] M. Schiffman,(2012) “The Missing Manual: CVRF 1.1,” ICASI Rep. [Online]. Available at: https://www.icasi.org/wp-content/uploads/2015/06/ICASI_CVRF1.1_White_Paper.pdf

[14] Hagen S. (2017,Sept) CSAF Common Vulnerability Reporting Framework (CVRF) Version 1.2. OASIS Committee Specification 01. [Online]. Available at: http://docs.oasis-open.org/csaf/csafcvrf/v1.2/csaf-cvrf-v1.2.html.

[15] Dshield / SANS Institute (2018) Internet Storm Center / DShield API. [Online]. Available at: https://dshield.org/api/

[16] EclecticIQ B.V. (2018) Cyber Threat Intelligence Analysis. [Online]. Available at: https://www.eclecticiq.com/

[17] EclecticIQ B.V. (2018) TAXIIStand - EclectiIQ TAXII Test Server. [Online]. Available at:

https://open.taxiistand.com/

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

[19] ENISA (2015). Standards and tools for exchange and processing of actionable information. [Online]. Available at: https://www.enisa.europa.eu/publications/standards-and-tools-for-exchange-and-processing-of-actionable-information

[20] ENISA (2015). Information sharing and common taxonomies between CSIRTs and Law Enforcement. [Online]. Available at: https://www.enisa.europa.eu/publications/information-sharing-and-common-taxonomies-between-csirts-and-law-enforcement

[21] ENISA (2016). Report on Cyber Security Information Sharing in the Energy Sector. [Online]. Available at: https://www.enisa.europa.eu/publications/information-sharing-in-the-energy-sector

[22] ENISA (2018). Information Sharing and Analysis Centres (ISACs): Cooperative models. [Online]. Available at: https://www.enisa.europa.eu/publications/information-sharing-and-analysis-center-isacs-cooperative-models

[23] G. Farnham, “Tools and Standards for Cyber Threat Intelligence Projects,” *SANS Institute InfoSec Reading Room*, October 14th 2013.

[24] Fail2ban.org. (2016), fail2ban. [Online]. Available at: https://www.fail2ban.org

[25] J. Friedman and M. Bouchard. (2015) *Definitive Guide to Cyber Threat Intelligence*, CyberEdge. [Online]. Available at: https://cryptome.org/2015/09/cti-guide.pdf

[26] FireEye Critical Infrastructure brief (2017) [Online]. Available at: https://www.fireeye.com/content/dam/fireeye-www/solutions/pdfs/pf/ms/sb-critical-infrastructure.pdf

[27] Greenbone Networks. (2018) The world's most advanced Open Source vulnerability scanner and manager. [Online]. Available at: http://www.openvas.org/

[28] B. Grobauer, T. Schrek, J. Goebel, J. Wallinger, and S. Berger. (2014) The MANTIS Framework: Cyber Threat Intelligence Management for CERTs. [Online]. Available at: https://www.first.org/resources/papers/conference2014/first_2014_-_grobauer-_bernd_-_mantis_20140624.pdf

[29] Hybrid Analysis (2018). Free malware analysis service for the community. [Online]. Available at: https://www.hybrid-analysis.com/

[30] J.L. Hernandez-Ardieta, J.E. Tapiador, and G. Suarez-Tangil, “Information sharing models for cooperative cyber defense,” in proc. of *5th International Conference on Cyber Conflict (CYCON 2013)*, pp. 63-90, 2013.

[31] EclecticIQ integration with Intel 471 Press Release. [Online]. Available at: https://www.eclecticiq.com/news/24-may-2016-eclecticiq-and-intel-471-partnership-threat-

intelligence

[32] ThreatStream and Intel 471 integration datasheet. [Online]. Available at: https://anomali.cdn.rackfoundry.net/files/data-sheets/intel471-datasheet.pdf

[33] Infoblox ActiveTrust datasheet. [Online]. Available at: https://www.infoblox.com/wp-content/uploads/infoblox-datasheet-infoblox-activetrust.pdf

[34] iSIGHT datasheet 2018. [Online]. Available at: https://www.fireeye.com/content/dam/fireeye-www/products/pdfs/pf/intel/ds-isight-threat-intelligence.pdf

[35] Jordan, B. (2016). Graphics, icons, and diagrams to support STIX 2. [Online]. Available at: https://github.com/freetaxii/stix2-graphics

[36] Kingfisher Operations. (2018) [Online]. Available at: https://kingfisherops.com/

[37] Malc0de Database. (2018) [Online]. Available at: http://malc0de.com/database/

[38] Mitre Corp., (2018). About MAEC. [Online]. Available at: http://maecproject.github.io/about-maec

[39] The MITRE Corporation. (2018) CVE - Common Vulnerabilities and Exposures. [Online]. Available at: http://cve.mitre.org/index.html

[40] The MITRE Corporation. (2018) Download CVE List - Download Formats. [Online]. Available at: https://cve.mitre.org/data/downloads/index.html

[41] Mitre Corp. (2018). MAEC Overview. [Online]. Available at: http://maecproject.github.io/documentation/overview/

[42] Mitre Corp. (2017). MAEC Vocabularies Specification [Online]. Available at: http://maecproject.github.io/releases/5.0/MAEC_Vocabularies_Specification.pdf

[43] Mitre Corp. (2018). VirusTotal to MAEC Utility. [Online]. Available at: https://github.com/MAECProject/vt-to-maec

[44] MANTIS Documentation. (2015) [Online]. Available at: https://media.readthedocs.org/pdf/django-mantis/latest/django-mantis.pdf

[45] NC4 / Soltra LLC. (2018). PickUpStix. [Online]. Available at: https://www.soltra.com/en/documentation/ctx-soltra-edge/connecting-to-pickupstix/

[46] Norse Appliance datasheet. (2015) [Online]. Available at: http://www.norse-corp.com/wp-content/uploads/2015/04/Norse_Appliance_DS_1019151.pdf

[47] Norse Appliance LE datasheet. (2015) [Online]. Available at: http://www.norse-corp.com/wp-content/uploads/2015/10/Appliance-LE_1019151.pdf

[48] The MITRE Corporation. (2018) CWE - Common Weakness Enumeration. [Online]. Available at: http://cwe.mitre.org/index.html

[49] Norse Intelligence Service datasheet. (2015) [Online]. Available at: http://www.norse-

corp.com/wp-content/uploads/2015/04/Norse-Intel-Service_DS_101915.pdf

[50] Norse Corporation Solutions Overview datasheet. (2015) [Online]. Available at: http://www.norsecorp.com/wp-content/uploads/2015/04/Corporate-Overview_101915.pdf

[51] NIST. (2013) Common Configuration Enumeration (CCE) Details - List [Online]. Available at: https://nvd.nist.gov/config/cce/index

[52] NIST. (2016) Guide to Cyber Threat Information Sharing. Special Publication 800-150.

[53] NIST. (2018) Information Technology Laboratory - National Vulnerability Database. [Online]. Available at: https://nvd.nist.gov/

[54] OASIS Open. (2018) Introduction to STIX. [Online]. Available at: https://oasis-open.github.io/cti-documentation/stix/intro

[55] OASIS Open. (2018). Introduction to TAXII. [Online]. Available at: https://oasis-open.github.io/cti-documentation/taxii/intro.html

[56] OASIS Open. (2017) STIX v2.0 - Part 2, STIX Objects. [Online]. Available at: http://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part2-stix-objects.html

[57] OASIS Open. (2017) STIX v2.0 - Part 5, STIX Patterning. [Online]. Available at: https://docs.oasis-open.org/cti/stix/v2.0/stix-v2.0-part5-stix-patterning.html

[58] OASIS Open. (2018) OASIS TC Open Repository: TAXII 2 Client Library Written in Python. [Online]. Available at: https://github.com/oasis-open/cti-taxii-client

[59] OASIS Open. (2018). OASIS TC Open Repository: TAXII 2 Server Library Written in Python. [Online]. Available at: https://github.com/oasis-open/cti-taxii-server

[60] Offensive Security. (2018) Exploit Database - Offensive Security’s Exploit Database Archive. [Online]. Available at: https://www.exploit-db.com/

[61] Offensive Security. (2018) Google Hacking Database (GHDB). [Online]. Available at: https://www.offensive-security.com/community-projects/google-hacking-database/

[62] Mitre Corp. (2006) Open Vulnerability and Assessment Language. [Online]. Available at: http://oval.mitre.org/documents/docs-06/an_introduction_to_the_oval_language.pdf

[63] P. Poputa-Clean, “Automated Defense Using Threat Intelligence to Augment Security,” *SANS Institute InfoSec Reading Room*, January 15th 2015.

[64] Rapid7 (2017) IoTSeeker - Locate connected IoT devices and check for default passwords. [Online]. Available at: https://information.rapid7.com/iotseeker.html

[65] S.J. Roberts and R. Brown, *Intelligence-Driven Incident Response*, O’Reilly, 2017.

[66] F. Skopik, *Collaborative Cyber Threat Intelligence: Detecting and Responding to Advanced Cyber Attacks at National Level*. CRC Press, 2018.

[67] SANS Institute. (2018) Internet Storm Center. [Online]. Available at: https://isc.sans.edu/

[68] Symantec DeepSight Intelligence datasheet. [Online]. Available at: https://www.symantec.com/content/dam/symantec/docs/data-sheets/deepsight-intelligence-ds.pdf

[69] Symantec DeepSight Portal White Paper. [Online]. Available at: https://www.symantec.com/content/en/us/enterprise/white_papers/b-21257288-deepsight-early-warn-en.us.pdf

[70] Soltra Edge brochure (2018) [Online]. Available at: https://www.soltra.com/media/resources/NC4_Soltra_Edge_Brochure.pdf

[71] Shodan. (2018) Leverage the Power of Shodan. [Online]. Available at: https://developer.shodan.io/

[72] Shodan. (2018) On-Demand Scanning. [Online]. Available at: https://help.shodan.io/the-basics/on-demand-scanning

[73] Sentinel IPS. (2018) CINS Army List. [Online]. Available at: http://www.ciarmy.com/#list

[74] Spamhaus. (2018) The Spamhaus Don't Route Or Peer Lists. [Online]. Available at: https://www.spamhaus.org/drop/

[75] ThreatConnect Intelligence Source (2018) datasheet. [Online]. Available at: https://www.threatconnect.com/wp-content/uploads/ThreatConnect-Intel-Source-Slick-Sheet.pdf

[76] Tenable. (2018) Nessus - Close Your Cyber Exposure Gap. [Online]. Available at: https://www.tenable.com/products

[77] TOR Project. (2018) Tor: Overview. [Online]. Available at: https://www.torproject.org/about/overview.html.en

[78] ThreatQ product brief. (2017) [Online]. Available at: https://www.threatq.com/documentation/threatq-threat-operations-and-management.pdf

[79] VeriSign iDefense and QualysGuard Vulnerability Management integration 2009 datasheet. [Online]. Available at: https://www.qualys.com/docs/iDefenseExcQualysGuard_Datasheet.pdf

[80] VeriSign iDefense and QualysGuard Vulnerability Management integration 2012 datasheet. [Online]. Available at: https://www.qualys.com/docs/VRSN_iDefense_QualysGuard_DS_20120604.pdf

[81] C. Sauerwein, C. Sillaber, A. Mussmann, and R. Breu, "Threat Intelligence Sharing Platforms: An Exploratory Study of Software Vendors and Research Perspectives," in Proc. der 13. Internationalen Tagung Wirtschaftsinformatik (WI 2017), pp 837-851, 2017.

[82] S. Barnum, Standardizing cyber threat intelligence information with the Structured Threat Information eXpression (STIXTM). Technical Report, MITRE Cooperation (2012)

[83] ENISA (2017). ENISA Threat Landscape Report 2017 -15 Top Cyber-Threats and Trends [Online]. Available at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-report-2017

11. Annex A. Information sensitivity and privacy

Considering privacy issues, a great challenge about threat information sharing is the probability of revealing PII. Awareness activities and researchers are supposed to ensure that those responsible for using threat information understand how to recognize it and protect PII. Sharing internal information may end up in discoloring PII to people who, should not typically have admission to such information. An organization should have procedures and information sharing policies to provide guidance for the treatment of PII. These procedures and policies should include steps to identify incident data that probable contain PII. Additionally, they should provide proper safeguards for detecting the risks concerning privacy which are associated with such sharing.

Organizations in order to identify and protect PII instead of using human-oriented methods are practically encouraged to use automated methods, which may contain field-level data validation. This is done by searching for PII using pattern matching techniques (e.g., regular expressions), and using techniques to anonymize and de-identify data containing PII. The degree of difficulty to achieve this varies depending on the sensitivity, complexity, structure of the information. In Table 33 we introduce selected types of CTI and provide examples of sensitive data and offers recommendations for the use of it.

Table 33. Handling of selected types of sensitive data

Type of CTI	Examples of sensitive data elements	NIST recommendations	Cyber-Trust approach
Network indicators	Any single network indicator can be sensitive, but network indicators in the aggregate are often more sensitive because they can reveal relationships between network entities. By studying these relationships it may be possible to infer the identity of users, gather information about the posture of devices, perform network reconnaissance, and characterize the security safeguards and tools that an organization uses.	Focus on the exchange of network indicators such as destination IP addresses associated with an actor's command and control infrastructure, malicious URLs/domains, and staging servers. Before sharing, anonymize or sanitize network indicators that contain IP or MAC addresses of target systems or addresses registered to your organization. Also anonymize or sanitize indicators that may reveal the structure of internal networks, or ports or protocols that identify particular products.	To be defined jointly with WP3

Type of CTI	Examples of sensitive data elements	NIST recommendations	Cyber-Trust approach
Packet capture (PCAP)	In addition to the network indicators previously discussed, unencrypted or decrypted packets may contain authentication credentials and sensitive organization information, such as PII, CUI or other types of sensitive information.	PCAP files can be challenging because network indicators may be present within both the packet header and the payload. For example, PCAP files may show protocols (e.g., DHCP, Address Resolution Protocol (ARP), File Transfer Protocol (FTP), DNS) and applications operating at multiple layers within the network stack. These protocols and applications generate network information that may be captured within PCAP files and may require sanitization or anonymization to prevent sensitive information leakage. Filter PCAP files before sharing by extracting only those packets that are related to the investigation of a specific incident or pattern of events: ● Related to a particular network conversation (i.e., exchange of information between specific IP addresses of interest); ● Occurring during a chosen time period; ● Destined for, or originating from, a specific port; or ● Use of a particular network protocol. Redact payload content that contains PII, CUI or other types of sensitive information that is not relevant for characterizing the incident or event of	To be defined jointly with WP3

Type of CTI	Examples of sensitive data elements	NIST recommendations	Cyber-Trust approach
		interest. When anonymizing or redacting network information, use a strategy that preserves enough information to support meaningful analysis of the resulting PCAP file contents.	
Network flow data	Network flow data contains information such as: ●Source IP address (i.e., the sender), ●Destination IP address (i.e., the recipient), ●Port and protocol information, ●Byte counts, and ●Timestamps. If not effectively anonymized, network flow data may make identification of specific users possible, provide insights into user behavior (e.g., web sites visited), expose application and service usage patterns, or reveal network routing information and data volumes.	Before sharing network flow data, organizations should consider redacting portions of session histories using cryptography-based, prefix-preserving, IP address anonymization techniques to prevent network identification or to conceal specific fields within the session trace (e.g., timestamps, ports, protocols, or byte counts). To gain the greatest value from the information, use a tool that transforms network flow data without breaking referential integrity. Network flow analysis and correlation operations often require that IP address replacement and transformation operations are performed consistently within and sometimes across multiple files. Anonymization techniques that do not use a consistent replacement strategy may reduce or eliminate the value of sharing this type of information.	To be defined jointly with WP3

Type of CTI	Examples of sensitive data elements	NIST recommendations	Cyber-Trust approach
Phishing email samples	Email headers may contain information such as: ●Mail agent IP addresses, ●Host or domain names, and ●Email addresses. An email message body may also contain PII, CUI, or other types of sensitive information.	Organizations should anonymize email samples and remove any sensitive information that is not necessary for describing an incident or event of interest.	To be defined jointly with WP3
System, network, and application log	Log files may contain PII, CUI or other types of sensitive information. Log data may reveal IP addresses, ports, protocols, services, and URLs, as well as connection strings, logon credentials, portions of financial transactions, or other activities captured in URL parameters.	Organizations should perform IP address, timestamp, port, and protocol anonymization and remove any sensitive information that is not necessary for describing an incident or event of interest. Before sharing log data, it may also be necessary to sanitize URLs that contain identifying information such as session or user identifiers. Application logs may require redaction and anonymizing operations that are specific to particular application log formats.	To be defined jointly with WP3
Malware Indicators and Samples	Although organizations are unlikely to encounter sensitive information in malware indicators or samples, sensitive information may be present depending on how targeted the malware is and what collection methods were used to gather a sample.	Organizations should remove PII, CUI, and other types of sensitive information that is not necessary for describing an incident or event of interest.	To be defined jointly with WP3

Type of CTI	Examples of sensitive data elements	NIST recommendations	Cyber-Trust approach
Community CTI	Indicators and observables obtained from community CTI sources may contain data that could, or could assist with the identification of a 3rd party	Organizations should ensure that data obtained from community sources should have any identifiable PII removed or anonymized, (e.g., contact email or author). General rules for data retention should be observed.	To be defined jointly with WP3
External CTI news feeds	News feed data contains PII	This data has been published and does not require sanitization.	To be defined jointly with WP3
External CTI advisories	Vulnerability advisories may contain contact or author details	This data may be deemed as published.	To be defined jointly with WP3
External CTI, search automation and dark web	Automated search activity, information from forums, paste bins, etc. may contain PII	Organizations should ensure that data obtained from external sources should have any identifiable PII removed or anonymized	To be defined jointly with WP3

12. Annex B. Sharing designations

Many methods exist to designate how shared CTI should be handled, e.g., by identifying unclassified or classified and sensitive information. The *traffic light protocol* (TLP) is such a method and specifies a set of restrictions, designated by a coloring scheme, that are applicable to a particular data record. This is also included here for completeness and is shown in Table 34 [52].

Table 34. The traffic light protocol

Color	When should it be used?	How may it be shared?
TLP:RED Not for disclosure, restricted to participants only	Sources may use TLP:RED when information cannot be effectively acted upon by additional parties, and could lead to impacts on a party's privacy, reputation, or operations if misused.	Recipients may not share TLP:RED information with any parties outside of the specific exchange, meeting, or conversation in which it was originally disclosed. In the context of a meeting, for example, TLP:RED information is limited to those present at the meeting. In most circumstances, TLP:RED should be exchanged verbally or in person.
TLP:AMBER Limited disclosure, restricted to participants' organizations	Sources may use TLP:AMBER when information requires support to be effectively acted upon, yet carries risks to privacy, reputation, or operations if shared outside of the organizations involved.	Recipients may only share TLP:AMBER information with members of their own organization, and with clients or customers who need to know the information to protect themselves or prevent further harm. Sources are at liberty to specify additional intended limits of the sharing: these must be adhered to
TLP:GREEN Limited disclosure, restricted to the community	Sources may use TLP:GREEN when information is useful for the awareness of all participating organizations as well as with peers within the broader community or sector.	Recipients may share TLP:GREEN information with peers and partner organizations within their sector or community, but not via publicly accessible channels. Information in this category can be circulated widely within a particular community. TLP:GREEN information may not be released outside of the community.
TLP:WHITE Disclosure is not limited	Sources may use TLP:WHITE when information carries minimal or no foreseeable risk of misuse, in accordance with applicable rules and procedures for public release. Subject to standard copyright rules.	TLP:WHITE information may be distributed without restriction.