SAFETY4RAILS

Framework and Methodology of Critical Components Based on OSINT

Deliverable 4.2

Lead Author: INNO

Contributors: TREE, IC, CS, UMH

Dissemination level: Public

Security Assessment Control: passed

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883532.

D4.2 FRAMEWORK AND METHODOLOGY OF CRITICAL COMPONENTS BASED ON OSINT		
Deliverable number:	D4.2	
Version:	1.0	
Delivery date:	31/03/2022	
Deliverable due date:	31/03/2022	
Dissemination level:	Public	
Nature:	Report	
Main authors:	Lesley Badii Marco Tiemann Ryan Faulkner	INNO INNO INNO
Contributor(s):	Alejandro Prada Nespral Tatiana Silva Diego Bernabé Zsuzsanna Keri Loca Molnar Eros Cazzato Uli Siebold Philippe Verdier	TREE TREE TREE CS CS IC IC IC
Internal reviewer(s):	Nacho Diaz Stephen Crabbe Uli Siebold Eros Cazzato Andreas Georgakopoulos Atta Badii Antonio De Santiago Laporte	UMH Fraunhofer IC IC WINGS UREAD MdM
External reviewer:	Reto Biedermann	IC

Document control			
Version	Date	Author(s)	Change(s)
0.1	20/12/2021	Marco Tiemann	ToC release
0.2	25/01/2022	Marco Tiemann	Initial contributions
0.3	31/01/2022	Marco Tiemann Lesley Badii Ryan Faulkner	INNO updates
0.4	02/02/2022	Alejandro Prada, Tatiana Silva Diego Bernabé	TREE updates regarding TISAIL
0.5	04/02/2022	Marco Tiemann	Revision Sections 1, 2, 5
0.6	25/02/2022	Uli Siebold Lesley Badii	Revision Section 3, 4
0.7	03/03/2022	Ryan Faulkner Luca Molnar	Revision Section 3, 4
0.8	14/03/2022	Marco Tiemann	Updates for review version
0.9	15/03/2022	Tatiana Silva Alejandro Prada	Revision Section 3.3.2, 4.2
1.0	31/03/2022	Marco Tiemann Tatiana Silva	Updates according to reviewer comments

DISCLAIMER AND COPYRIGHT

ABOUT SAFETY4RAILS

SAFETY4RAILS is the acronym for the innovation project: **Data-based analysis for SAFETY and security protection FOR detection, prevention, mitigation and response in trans-modal metro and RAILway networkS**. Railways and Metros are safe, efficient, reliable and environmentally friendly mass carriers, and they are becoming even more important means of transportation given the need to address climate change. However, being such critical infrastructures turns metro and railway operators as well as related intermodal transport operators into attractive targets for cyber and/or physical attacks.**The SAFETY4RAILS project delivers methods and systems to increase the safety and recovery of track-based inter-city railway and intra-city metro transportation.** It addresses both cyber-only attacks (such as impact from WannaCry infections), physical-only attacks (such as the Madrid commuter trains bombing in 2014) and combined cyber-physical attacks, which are important emerging scenarios given increasing IoT infrastructure integration.

SAFETY4RAILS concentrates on rush hour rail transport scenarios where many passengers are using metros and railways to commute to work or attend mass events (e.g. large multi-venue sporting events such as the Olympics). When an incident occurs during heavy usage, metro and railway operators have to consider many aspects to ensure passenger safety and security, e.g. carry out a threat analysis, maintain situation awareness, establish crisis communication and response, and they must ensure that mitigation steps are taken and communicated to travellers and other users. **SAFETY4RAILS will improve the handling of such events through a holistic approach.** It will analyse the cyber-physical resilience of metro and railway systems and deliver mitigation strategies for an efficient response, and, in order to remain secure given everchanging novel emerging risks, it will facilitate continuous adaptation of the SAFETY4RAILS solution; this will be validated by two rail transport operators and the results will support the re-design of the final prototype.

TABLE OF CONTENTS

List of tables

List of figures

Executive summary

This deliverable report, D4.2 "Framework and Methodology of Critical Components Based on OSINT", reports on the activities that have been carried out in Task 4.2 "OSINT Technologies for Cyber-Physical Intelligence (SCADA)". The report expands on initial reporting undertaken in the task that has been reported in the previous deliverable D4.3 "Cyber-Physical Threat Detection with Capabilities Matrix Intelligence" and incorporates progress made during the development activities undertaken in the task since the preparation of that report. Where applicable, content from deliverable D4.3 that remains valid has been retained in this deliverable in order to create a single resource documenting the work undertaken in Task 4.2.

The deliverable introduces the concept of Open Source Intelligence (OSINT) in Section 2. Section 3 describes the design and overall implementation of the OSINT system developed in the project. Section 4 documents the data modelling, data sources and data acquisition that have been undertaken and integrated in T4.2. Section 5 concludes the deliverable with a summary of activities, a summary of the task activities and an outlook towards next steps both within and beyond the scope of SAFETY4RAILS.

1. Introduction

This introductory section provides an overview of the activities in Task 4.2 on which the deliverable reports and outlines the structure and contents of the deliverable.

1.1 Activity Overview

This document reports on the activities undertaken in Task 4.2 “OSINT Technologies for Cyber-Physical Intelligence (SCADA)”. We first introduce the activities, goals and means that were important for this task in order to provide the necessary understanding for the remainder of this deliverable.

The work reported on in this deliverable has been concerned with retrieving, structuring, analysing and making available open source intelligence (OSINT). Open source intelligence can generally be described as analytics using openly available data (see Section 2.1 for a more differentiated definition). The main aims of this task were to gather, process and make accessible open source intelligence that relates to cyber-physical vulnerabilities and threats of concern within the scope of the SAFETY4RAILS project and of course to those who may benefit from these activities beyond the project consortium. Achieving these aims involved the following main activities:

1. Specification and implementation of a data model that suitably represented the application domain in terms of components, their potential vulnerabilities and past and present threats relevant to the components under consideration. The data model can represent cyber, physical and cyber-physical data in line with the overall ambitions of the SAFETY4RAILS project.
2. Integration of an open source intelligence repository implementation using the developed data model; the repository implementation supports typical database functionalities and also readily integrates into the overall SAFETY4RAILS system infrastructure and provides means for retrieving data from the repository as well as for communicating newly identified information via the distributed messaging system used for inter-component communication between the project’s software components.
3. Development of open source data retrieval components that collect relevant open source intelligence data from web sources. These integrate commonly used sources such as malware repositories and threat intelligence feeds that are relevant for the application domain as well as less formal sources including social media feeds both for cyber vulnerabilities and physical threats[2]. The gathered data is processed so that the extracted information can be stored in the database using the data model developed for the project. The data gathering process has been automated as far as possible while maintaining an approach that allows administrators to configure the risks they would like to monitor via the platform.
4. Development of alerting and simple statistics processing functionalities in order to create necessary alerting functionalities for vulnerable components and in order to generate overview summaries of key risks in the application domain for more strategic review and reaction purposes.
5. Populating the database with component, vulnerability and threat data for technical testing as well as for use in the SAFETY4RAILS demonstration and simulation activities.

Two deliverable documents are part of the work in the task, this deliverable and the deliverable D4.3 “Cyber-Physical Threat Detection With Capabilities Matrix Intelligence”, which was delivered in project month 8. Furthermore, the task contributes to the overall data modelling, software development and evaluation activities in the project; additional contributions relevant to the work of Task 4.2 are available in the relevant deliverables of the respectively responsible work packages.

[2] In the project, we use social media sources operated by organisations concerned with publishing information on vulnerabilities and additionally sources created by us that simulate private user social media feeds. With these approaches we aim to avoid gathering any real personal data from social media sources.

1.2 Organisation of this Deliverable

The main aims of this deliverable are to document the methods for cyber-physical threat detection using open source intelligence that the task has developed throughout the overall task duration. The deliverable furthermore summarises the contributions and focus areas of the individual components and other contributions in the task through a capability matrix for open source intelligence that was also presented as part of D4.3 in order to provide the reader with a single source of the final information on the output of Task 4.2. Some content that was presented in D4.3 has been carried over in this deliverable release in order to present a single source of information for Task 4.2. The introductory text at the beginning of each section will specify clearly which parts of each section have carried over substantial amounts of content from D4.3 for clarity.

The remainder of this deliverable is organised as follows:

- Section 2 “Open Source Intelligence” introduces the concept of open source intelligence generally as well as specifically in the context of SAFETY4RAILS. The section presents the main functionalities provided via open source intelligence in SAFETY4RAILS.
- Section 3 “Open Source Intelligence System Design” describes the design implemented for the OSINT system in SAFETY4RAILS. It also introduces the relevant background information on tools that are included as part of the design and the developed system.
- Section 4 “Data Model, Data Sources & Data Acquisition” is concerned with a description of the data model used in Task 4.2 (including the relation to data models developed in other tasks of the project), with the identification of data sources for the task and with the process of data acquisition used during the execution of the task activities.
- Section 5 “Conclusion” summarises the contents of this deliverable and looks forward towards the follow-up activities during the remainder of the project as well as after the completion of the project. The section concludes with a matrix that relates the task activities and outcomes to the functionalities discussed in Section 2 of the deliverable.

A list of references, an annex with glossary and acronym explanations and an annex describing the toy example data model used during technical development are appended at the end of the document.

2. Open Source Intelligence

This section introduces the field of Open Source Intelligence in general as well as in terms of how it is applied within SAFETY4RAILS. Based hereon, the main functionalities of open source intelligence for SAFETY4RAILS are specified.

Please note that this section largely reproduces content initially introduced in deliverable D4.3 in Section 2.1 and Section 2.2 due to the nature of the content provided there. Section 2.3 has been updated with a review of the system requirements set out for the OSINT subsystem (including TISAIL-specific requirements) regarding to which extend they have been realised through the work undertaken in Task 4.2.

2.1 Definition

The term “open source intelligence” (OSINT) originated and is frequently used in the intelligence community (including military intelligence). In this context, (1) defines it as “unclassified information that has been deliberately discovered, discriminated, distilled and disseminated to a select audience in order to address a specific question.” Further definitions that also focus on a definition from a military perspective as can be found for instance in (2) generally define OSINT similarly with some variations in terms of the scope of what is included in under “open source” (publicly available, legally available, unclassified information, etc.). (3) more generally defines OSINT as “all information that can be derived from overt collection”.

As the described activities are not undertaken in a military context, the following more generic definition in (2) is useful:

> “Open-source intelligence (OSINT) is a multi-factor […] methodology for collecting, analyzing and making decisions about data accessible in publicly available sources to be used in an intelligence context. In the intelligence community, the term ‘open’ refers to overt, publicly available sources (as opposed to covert or clandestine sources).”

The term “open source” in the context of OSINT is in the remainder used to denote publicly available sources of potential intelligence value, and “intelligence” is defined as the process of collecting, analysing and using the collected information in order to generate useful information for use by others.

2.2 Open Source Intelligence in SAFETY4RAILS

What is the rationale for and what is the purpose of having open source intelligence as a part of the activities in SAFETY4RAILS?

Based on the general definition above, OSINT in SAFETY4RAILS should be used to gather and evaluate “open source” data in order to support the goals and functionalities envisioned for the project, including identifying weaknesses in live systems, the automation of systems dedicated to monitoring cyber vulnerabilities and supporting the prevention of risks and threats as identified in deliverable D2.1 “Grid Analysis of End-User Needs and Workshop Minutes”. One key aspect of the project is the ambition to cover a wide range of cyber, physical and combined cyber-physical systems in order to increase their security as well as overall rail infrastructure safety. Open source intelligence can provide valuable information in these areas by identifying vulnerabilities, threats and, where feasible from open source data, specific information on vulnerabilities or suspicious behaviour identified in publicly accessible systems and services.

It is useful to take note of current typical real-world usage of open source intelligence, particularly mediated by technological solutions such as the ones proposed in SAFETY4RAILS, separately for cyber, physical and combined cyber-physical security in general as well as in the railway domain. The following table summarises the view of the task participants based on their expertise in the field.

TABLE 1: USE OF OSINT IN CYBER, PHYSICAL AND CYBER-PHYSICAL SCENARIOS

Domain	*Current Usage in Industry Overall*	*Current Usage by Rail Operators*
Cyber security	OSINT is used fairly widely, sources such as malware repositories or threat data feeds are available for general IT infrastructure and also for Supervisory Control and Data Acquisition (SCADA) devices of major manufacturers; standard data sources, organisational structures and software tools for managing threat repositories are available	Railway use of OSINT in the cyber security domain is generally comparable with OSINT use in industrial sectors; standard not domain-specific threat repository systems are used by some operators; security matters may be outsourced to third parties
Physical security	OSINT is not used widely in order to identify physical threats such as for instance natural hazards from open source data outside of the security domain	Little to no use of OSINT in order to identify physical threats in the railway domain
Cyber-physical security	OSINT usually only used by more sophisticated operators in the security domain	No use of OSINT in the cyber-physical domain; cyber and physical risks are not usually managed by the same departments or responder groups

Generally, OSINT tools used in the railway domain focus on cyber security and are not specialised for use in the railway domain. The goal of developing open source intelligence tools in SAFETY4RAILS has therefore included a) the development of specialised solutions for the railway domain in terms of data gathering and processing, b) extending the coverage of OSINT technologies to also cover physical and where practical cyber-physical security and c) automating data processing for OSINT acquisition and processing as far as possible given the quality of real-world data sources available. These specific goals have been defined in more detail as functionalities to be provided by OSINT for use in SAFETY4RAILS and the implementation carried out in T4.2 has focused on realising the specified functionalities.

2.3 Requirements Addressed

In order to be useful for the type of end user organisations addressed by SAFETY4RAILS, an OSINT system must target the specific potential sources of risks, such as specific deployed component models that are used by an end user organisation, in order to ensure that intelligence is gathered for the relevant infrastructure and environment. The availability of an inventory of systems and components to be included into OSINT processes is therefore a prerequisite for the targeted use of OSINT in SAFETY4RAILS. These data are used to customise search processes and filter out irrelevant data from OSINT analyses; the OSINT system should provide an easy to use API to add and update this information as changes are made to the railway operator infrastructure[4].

The data processing functionalities required to be provided via OSINT are derived from the requirements specified for the OSINT functionalities in SAFETY4RAILS, set out in deliverable D1.4. For specification purposes, we organised the resulting functionalities by first specifying a processing pipeline for the OSINT system in SAFETY4RAILS and then describing the functionalities to be provided at each stage of the

[4] Section 4.1 describes the integration of such data from a data model perspective and Annex II describes the data used for demonstration and validation activities during the project. The system API depicted in Figure 3 exposes methods to create, read, update and delete relevant entries in the MISP database.

processing pipeline. At the end of this subsection, we revisit the requirements defined in D1.4 and review to which extend the specified requirements have been met.

FIGURE 1: SAFETY4RAILS OSINT PROCESSING PIPELINE

This processing pipeline is applied for each data processing activity. Processing activities are recurring activities that are carried out in regular intervals upon being triggered by notifications of available data updates or activated by changes in the infrastructure data entered by railway operators through the MISP or TISAIL user interfaces.

We describe each of the processing steps together with the associated functionalities that are provided in the table below.

TABLE 2: FUNCTIONALITIES ORGANISED BY PROCESSING PIPELINE STEPS

Processing Step	*Description (in italics) and functionalities*
Data Acquisition	*This processing step encompasses all activities that involve retrieval of OSINT data (push and pull)*
	FUNC-DA-01: Retrieval from typical cyber security data sources such as specialised search engines and data feed; data sources and retrieval from searches are preselected from data sources considered relevant for the domain; potentially retrieval from social media sources such as cyber security Twitter feeds
	FUNC-DA-02: Search for and identification of vulnerable and potentially compromised devices that are accessible on the Internet and relevant for the application domain
	FUNC-DA-03: Retrieval from relevant physical security data sources including specialised providers, general news feeds and potentially social media sources such as Twitter posts related to relevant tags
Pre-Processing and Analytics	*This processing step encompasses necessary activities for parsing the acquired data and for attempting to extract relevant information from any data that may not be sufficiently well structured for relatively simple parsing*
	FUNC-PP-01: Parsing of standard format data feeds and standard format search query results from specialised data sources
	FUNC-PP-02: Analytics and entity extraction from semi-structured and unstructured data sources such as free text and social media communications in order to identify potentially relevant data
Storage and Representation	*This processing step involves storing the processed data in a database while adhering to the database structure*

	FUNC-SR-01: Operation of a suitable database management system or similar infrastructure that provides typical utility functionalities
Data Set Analytics	*This processing step involves all analytics activities that are carried out on the database instead of a single data point prior to addition to the database*
	FUNC-DS-01: Analysis of newly added data points in order to identify new vulnerabilities or threats that are identified based on the newly added data point
	FUNC-DS-02: Generation of statistics in order to identify trends and rankings of threats and vulnerabilities
Data Access and Messaging	*This processing step consists of responding to data retrieval requests via an API and of messaging updates to components that expect notifications of particular state changes in the database*
	FUNC-DM-01: Exposing data access functionalities to authorised system components within SAFETY4RAIL
	FUNC-DM-02: Communication of data updates to components that require data to be pushed to them

These functionalities are related to the specific implementation tasks in which they have been implemented that are described in Section 3 of this deliverable. The functionalities are also relevant in the selection of data sources that are described later in this deliverable in order to specify how the defined functionalities are realised in the project.

The system requirements specified for the OSINT subsystem in SAFETY4RAILS and for TISAIL as a component of the OSINT subsystem describe the required functionalities derived from user requirements, project scenarios and the goals of the project as set out in the Description of Action of the project. Table 3 lists the system requirements set out in D1.4 and documents to whether and to which extent they have been addressed during the work in T4.2. For brevity, please refer to the full requirements specifications in D1.4 for details on the requirement wording.

TABLE 3: IMPLEMENTATION STATUS OF SYSTEM REQUIREMENTS FOR OSINT AND TISAIL

Req. ID	*Short name*	*Priority*	*Fulfilment*	*Explanation*
OSINT_1	Data acquisition of OSINT	Essential	100%	The OSINT subsystem can retrieve railway domain cyber and physical security OSINT data and identify vulnerable devices exposed to the Internet
OSINT_2	Pre-Processing and Analytics	Essential	100%	The OSINT subsystem can parse OSINT threat data including from data feeds and specialised data sources as well as extract entities from semi- and unstructured data sources (RSS feeds and social media messages)
OSINT_3	Storage and representation	Essential	100%	A dedicated MISP instance stores and represents OSINT data
OSINT_4	Data set analytics	Conditional	90%	The OSINT subsystem generates statistics to highlight prevalent as well as recently emerging main

				threat types and generates daily, weekly, monthly and lifetime statistics of threat types; new vulnerabilities, threats and events are communication via DMS events; MISP rule system provides mechanism for rule evaluation, so no custom rule engine was implemented (which had originally been envisioned)
OSINT_5	Data access and mechanism	Essential	100%	MISP API provides full access to threat data, OSINT subsystem is connected to Apache Kafka and communicates threat data via the distributed messaging system
TISAIL_1	Detection of cyber-threats related to the railway sector: malware	Essential	100%	TISAIL searches for relevant malware as specified in the requirement details
TISAIL_2	Detection of cyber-threats related to the railway sector: Internet-Exposed Assets and credential leaks	Essential	100%	TISAIL can detect specified threats as specified in the requirement details
TILSAIL_3	Detection of cyber-threats related to the railway sector: Threat Intel feeds and Social Media	Essential	100%	TISAIL can detect cyber-threats from threat intel feeds and social media as specified in the requirement details
TISAIL_4	Detection of cyber-threats related to the railway sector: Vulnerabilities	Essential	100%	TISAIL can detect cyber vulnerabilities as specified in the requirement details
TISAIL_5	Detection of cyber-threats related to the railway sector: Spear Phishing	Optional	100%	TISAIL has implemented functionalities to detect spear phishing approaches with a list of candidate spear phishing URLs that may be candidates for such campaigns in the railway sector
TISAIL_6	Integrate alerts related to cyber-threats in the railway sector with a MISP repository	Essential	100%	TISAIL is integrated with the OSINT MISP repository and send alerts to that MISP instance for storage and representation
TISAIL_7	Use a railway threat taxonomy on TISAIL	Optional	100%	TISAIL supports a railway taxonomy from X2Rail and applies tag information so that data gathered can be meaningfully processed and represented via RAM[2]
TISAIL_8	Conformity with overarching and S4RIS platform specific requirements included in Section 2.2 [of D1.4]	Essential	100%	TISAIL conforms with the requirements set out in Section 2.2 of D1.4 (n.b. cost-benefit balance was not assessed as part of the work in T4.2)

Section 3 presents an overview of how the OSINT subsystem has been implemented and configured in order to fulfil these requirements at a technical level.

3. OSINT System Implementation

This section describes the design and implementation of the software system that forms part of the overall solution created in SAFETY4RAILS. First, relevant development background that has influenced the system design is introduced. Then, the components that form part of the system design and implementation are introduced and individually described at implementation level. The system architecture is described including the integration into the overall SAFETY4RAILS system.

Section 3.1 is largely reproduced from D4.3 as it introduces general concepts. The remaining subsections of this section present the overview of components and architecture and present information on how the components presented have been implemented and configured and how the OSINT system has been deployed for use in SAFETY4RAILS.

3.1 Background

One important way in which the SAFETY4RAILS project accomplishes its objectives is by leveraging existing specialised software solutions and integrating them into an overarching system that is tailored to the target application domain. The majority of SAFETY4RAILS technical tasks have hence involved the integration of and extension to and/or customisation of an existing software solution. Task T4.2 includes the integration of TISAIL, a threat intelligence solution for the railway sector developed by TREE Technology. We introduce TISAIL here as background to help understand and motivate the overall T4.2 system design. Before that, we briefly introduce the Malware Information Sharing Platform (MISP[5]) that is used both in TISAIL and in the OSINT system design for the same reasons.

3.1.1 Malware Information Sharing Platform

The Malware Information Sharing Platform (MISP) is an open source, community-driven platform for the exchange and sharing of threat intelligence and Indicators of Compromise (IoCs) about targeted malware and attacks. The development of the platform was supported by NATO and the platform is currently widely used as a means to store and exchange threat data within trusted groups, especially in the cyber security domain. Beyond its applicability in cyber security, MISP implements a number of useful features for the scope of SAFETY4RAILS out of the box, including the following[6]:

- A widely used threat data model that is expressed in JSON format and can be customised for domains beyond the typical cyber security application domain of MISP;
- Integration of a database management system, a standard REST API, data synchronisation mechanisms as well as the ability to export threat data for direct import into commonly used intrusion detection systems (IDS) and custom tools (in the following formats: STIX, OpenIOC, plain text, CSV, XML, JSON);
- A correlation finding component that can automatically identify relationships between attributes and indicators from malware or data submitted to the MISP instance. This component was also extended with an advanced visualisation tool in the most recent rounds of development, to support analysis work within MISP. The cyber threat analysis process is further supported by the availability of a MITRE ATT&CK matrix integration;
- A flexible tool for importing and integrating MISP and OSINT data feeds using standard data formats with typically used feeds being available as part of a default installation. Further feeds can be integrated for free or on a subscription[7] basis, mostly offered by nation-level, CERT/CSIRT organisations or private cyber threat intelligence providers;

[5] More information on MISP is available on the project website at https://www.misp-project.org (accessed 29.03.2022).
[6] Please see https://www.misp-project.org/features.html (accessed 29.03.2022) for a more complete listing of MISP platform features.
[7] Subscription services may or may not be considered to be OSINT depending on the definition applied. Our aim in SAFETY4RAILS is to avoid subscription services within the context of the project, but to support their use in general for post-project deployment scenarios. No subscription services are used for use in SAFETY4RAILS.

- Well-defined expansion points and available expansion modules that enable customisation of MISP functionalities and for the easy integration and use of expansion modules that are made publicly available by the large user community.

MISP development has been co-funded by the European Union through the Connecting Europe facility. The core MISP platform and already developed modules are released under a GPLv3 licence (4), which means that modifications to the core software or modules must also be published as open source using the same licence. This does not apply to customised developments that use MISP APIs and similar software development interfaces[8].

3.1.2 TISAIL

TISAIL is a threat intelligence platform for the railway sector developed by SAFETY4RAILS project partner TREE Technology. Threat intelligence is here defined as follows (5):

> "Threat intelligence (TI) is evidence-based knowledge (e.g. context, mechanisms, indicators, implications and action-oriented advice) about existing or emerging menaces or hazards to assets. It can be used to inform decisions regarding the subject's response to that menace or hazard." (Gartner)

TISAIL incorporates three different stages as part of the threat intelligence process:

1. Using automated processes for discovering potential threats using threat intelligence feeds, malware repositories, vulnerability reports and detection rules.
2. Carrying out malware analysis processes.
3. Extracting Indicators of Compromise (IoC) and enriching the gathered information in order to generate threat data and notifications for use by other systems.

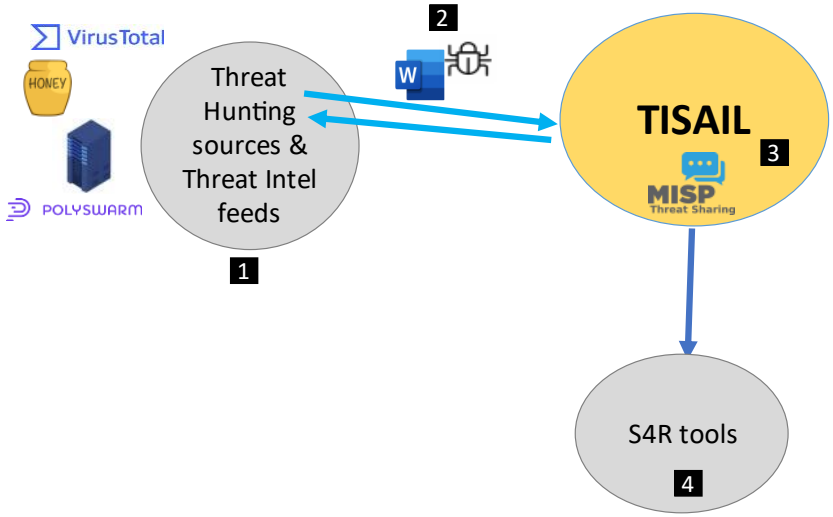
TISAIL uses the MISP (6) data model for data representation and uses MISP internally for data storage. A domain-specific taxonomy for threats is used to help decision makers identify and classify threats and take suitable actions more quickly. Figure 2 illustrates the general operation of TISAIL as part of SAFETY4RAILS.

The format of the TISAIL alerts is based on the MISP standard format that is used to exchange threat information among MISP instances. The MISP format is typically represented in JSON format when not stored directly in a database. It includes different standards for facilitating the representation of technical and non-technical information about malware, cyber-attacks and threat actors. MISP uses core and extension data models to support different application areas; TISAIL uses the MISP core format, the MISP object template and the MISP taxonomy format for communicating information. Details of each standard are explained in Section 4.1.

The communicated threat information may include IoCs, malicious file indicators, malware signatures or event information about a threat actor, such as its Tactics, Techniques and Procedures (TTPs).

TISAIL is a proprietary solution developed by TREE Technology, and the TISAIL infrastructure operates in a secured Amazon Web Services (AWS) environment in order to protect the gathered data (see also Section 3.3.2). This is relevant for the design of the overall solution in SAFETY4RAILS because other partners in the project will not be able to directly access TISAIL, and they will not be able to directly integrate software components with TISAIL itself. Since TISAIL uses the MISP data format and MISP repository as part of the overall solution, a suitable solution to address this limitation is to deploy a SAFETY4RAILS MISP instance with which TISAIL can communicate natively and with little additional effort.

[8] See https://www.misp-project.org/license/ (accessed 29.03.2022) for the MISP description of the applicability of the GPL license for MISP.



1. TISAIL discovers potential threats using Threat Intel feeds, Vulnerability reports and detection rules.
2. The information is analysed in a malware lab (TREE)
3. After analysing the threats, extracting some Indicators of Compromise and enriching the information, an alert is created.
4. The alerts are sent to other SAFETY4RAILS tools, so the end-user can see the alerts through other S4R tools (such as RAM2).

FIGURE 2: TISAIL PROCESSING OVERVIEW

3.2 Components

This subsection describes the components that have been identified as necessary and implemented for the SAFETY4RAILS OSINT system. The functionalities are briefly described in the tables included in each of the following subsubsections. Implementations of the components are then described grouping the implemented components into the unit level at which they were implemented. The subsubsections are organised analogously to the data processing pipeline structure shown in Section 2. As mentioned in the introduction for the section, the table descriptions of the necessary components were originally created for D4.3 and are reproduced here for completeness and to create a single source of reference.

3.2.1 Data Acquisition

Data acquisition components acquire data from open sources that may be relevant in order to provide intelligence concerning vulnerabilities, threats and general risks that are relevant to the application domain and can be gathered from open sources. Table 4 lists the currently specified data acquisition components. Please note that the term “component” is used in a wide sense and also includes specific system configurations (for instance for the retrieval of threat intelligence feed data).

TABLE 4: DATA ACQUISITION COMPONENTS

ID	*Title*	*Description*
DA-TIS-01 (TISAIL)	**Internet-exposed asset crawlers**	Implementation of a set of crawlers that search for IT/OT assets/components that are exposed on the Internet and gather information about them for correlation with a list of railway keywords as well as a list of products used by railway companies. In case of a match, the exposed asset is stored and an alert is created.
DA-TIS-02 (TISAIL)	**ICS Vulnerability Crawlers**	Implementation of a set of crawlers that search for known vulnerabilities and exploit ICS assets/components that have been defined as relevant in the railway domain

DA-TIS-03 (TISAIL)	**Malware repository crawlers with Yara rules**	Use of Yara[10] rules for tracking malware families targeting ICS/SCADA Systems. The rules will be deployed to query against malware databases such as the openly available malware sample repository MalwareBazaar[11].
DA-TIS-04 (TISAIL)	**Malware social media crawlers**	Development of a set of crawlers that retrieve IoC, TTPs and context about malware and Threat actors from binaries from Threat Intel feeds and social media sources (e.g., Twitter)
DA-TIS-05 (TISAIL)	**Phishing campaign monitoring crawlers**	Development of a crawler that monitors domain names for alterations of railway domain names, detecting potential phishing campaigns
DA-TIS-06 (TISAIL)	**Threat intel feed selection and configuration**	Selection and configuration of relevant threat intelligence feeds and social media that provide relevant information for processing in TISAIL
DA-MSP-01	**Threat intel feed selection and configuration for cyber threats**	Selection and configuration of domain-relevant threat intelligence feeds that relate to cyber security threats relevant to the railway domain and the particular system configurations of concern
DA-MSP-02	**Structured data source integration for physical threat data**	Development of crawlers that retrieve data concerning physical threat detection from structured data sources in well-defined scenarios (natural hazards, public safety emergencies) where those data source are available as open sources
DA-MSP-03	**Social media data source integration for physical threat data**	Development of crawlers that retrieve data concerning physical threat detection from social media sources (either real social media sources or simulated data sources)

Identifiers for data acquisition components contain “TIS” when they have been implemented as part of TISAIL in order to differentiate them from components that are not implemented directly as part of that system. Identifiers for components contain “MSP” when they are implemented as part of the project MISP instance. The descriptions of the implementations in this subsection are analogously organised by implementations undertaken within the TISAIL system and implementation undertaken using the SAFETY4RAILS MISP instance. Configuration details of relevant data sources are discussed in Section 4.2 and are not discussed here (concerns DA-TIS-06, DA-MSP-01 and parts of DA-MSP-02 and DA-MSP-03).

TISAIL Data Acquisition Implementation

The components DA-TIS-01, DA-TIS-02, DA-TIS-03, DA-TIS-04 and DA-TIS-05 were implemented in the TISAIL configuration deployed for SAFETY4RAILS.

[10] Yara is primarily a tool for identifying and classifying malware samples, see http://virustotal.github.io/yara/ (accessed 29.03.2022) for a general description of Yara, see https://blog.malwarebytes.com/security-world/technology/2017/09/explained-yara-rules/ (accessed 29.03.2022) for a description of the Yara rules format.

[11] https://bazaar.abuse.ch/ (accessed 29.03.2022)

Proprietary TISAIL components for identifying internet-exposed assets were configured for use in the SAFETY4RAILS simulation exercises using device type information provided for use in the exercises. The same information was employed in order to formulate relevant data acquisition rules for vulnerability and exploit scanning as well as for retrieving relevant data from malware repositories and malware-related social media sources. Railway and transport-related URLs were used in order to implement realistic phishing campaign monitoring functionalities through TISAIL.

DA-TIS-01 EXPOSED ASSETS

This tool allows users to gather IT/OT exposed assets for threat intelligence analysis using the ZoomEye[12] and Shodan[13] search engines. The tool uses the official Python Shodan[14] and ZoomEye[15] libraries.

FIGURE 3: EXPOSED ASSEST SOURCE CODE

The tool uses a list of search terms that is be used for searching in both search engines. In this case, the results provided focus on ICS/SCADA and CCTV/IP-Cameras as well as other IT devices that might be used by adversaries at early stages of the kill chain such as remote desktop systems or virtual network computing.

[12] Zoomeye, https://www.zoomeye.org/ (accessed 31.03.2022)

[13] Shodan, https://www.shodan.io/ (accessed 31.03.2022)

[14] Shodan, https://github.com/achillean/shodan-python (accessed 31.03.2022)

[15] ZoomEye, https://github.com/knownsec/ZoomEye-python (accessed, 31.03.2022)

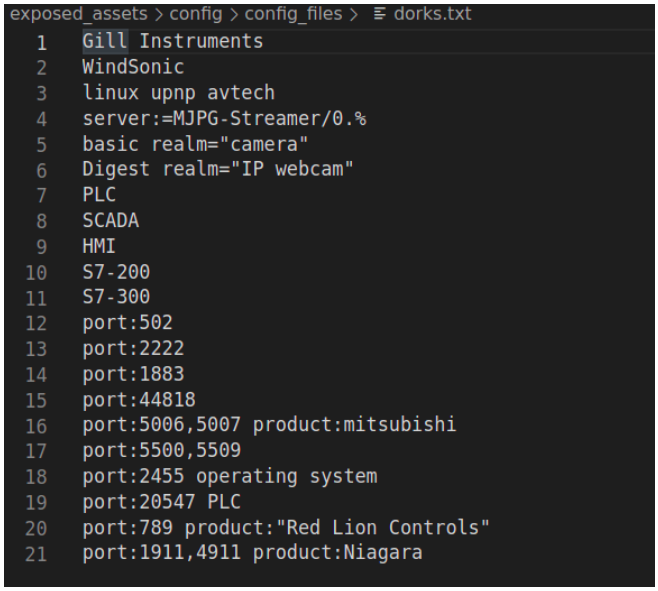


FIGURE 4: CRAWLER DEVELOPED FOR EXPOSED ASSETS

After gathering results from the different search engines, the tool has the functionality of correlating the banner of each host with a list of keywords. In case of a match, alerts will be sent to the project MISP instance.

DA-TIS-02 ICS VULNERABILITY CRAWLER

The aim of this tool is to provide a simple and automated way of gathering alerts about vulnerabilities and threats regarding ICS/SCADA. The tool uses the feedparser[16] Python library for consuming RSS feeds published by CISA, which regularly publishes potentially relevant alerts. The alerts gathered are correlated to a keywords list to find matches for the vulnerabilities that are being monitored. If any of the alerts contain one or more keywords stored in the software.txt file, the alerts will be sent to the configured project MISP instance.

The tool gathers information from the following sources within the CISA National Cyber Awareness System (see Section 4.2.1):

- Bulletins: Weekly summaries of new vulnerabilities (including patch information if available)
- Advisories: Timely information about current security issues, vulnerabilities and exploits

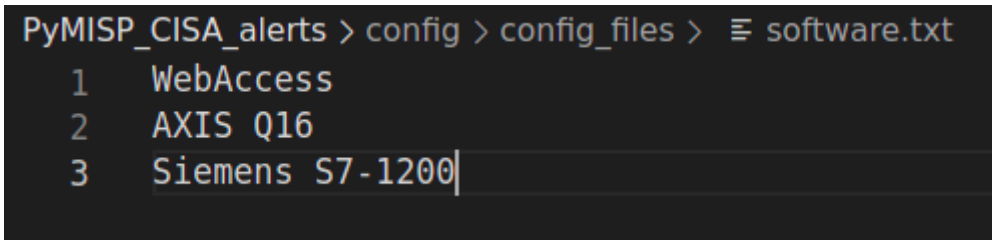


FIGURE 5: SOFTWARE MONITORED

In MISP, the alerts containing any of the keywords stored in the software.txt file are sent to the configured MISP instance. The events created contain the tag "vulnerability". An example of alert published on MISP is available in the figure below:

[16] feedparser, https://pypi.org/project/feedparser/ (accessed 31.03.2022)

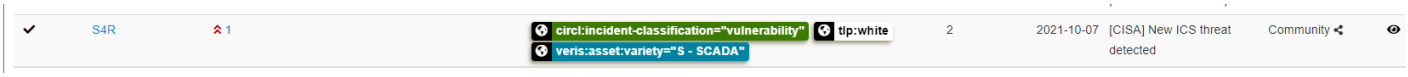


FIGURE 6: VULNERABILITY PUBLISHED IN THES4R MISP

DA-TIS-03 MALWARE REPOSITORY CRAWLERS WITH YARA RULES

The aim of this module is to develop and to use Yara rules provided by the Information Security Community to find malware samples related to ICS/SCADA devices. These rules will be used for trying to find malware samples in popular malware repositories (e.g. Hybrid Analysis, Malware Bazaar or Any.run, see Section 4.2.1). Figure 7 depicts an example of such a Yara rule.

```
rule BlackEnergy3

{

    strings:

        $a1 = "MCSF_Config" ascii

        $a2 = "NTUSER.LOG" ascii

        $a3 = "ldplg" ascii

        $a4 = "unlplg" ascii

        $a5 = "getp" ascii

        $a6 = "getpd" ascii

        $a7 = "CSTR" ascii

        $a8 = "FONTCACHE.DAT" ascii

    condition:

        4 of them

}
```

FIGURE 7: BLACK ENERGY YARA RULE

DA-TIS-04 MALWARE SOCIAL MEDIA CRAWLERS

This script allows gathering Indicator of Compromise (IoCs) from Twitter, looking for keywords (see examples in Figure 8), and send them to MISP for Threat Intelligence analysis (see Figure 9 and Figure 10).

```
ioc_tweet > config > filters.txt
1   APT
2   ICS
3   SCADA
4   RDP
5   exploit
6   0day
7   xloader
8   CVE-
```

FIGURE 8: CRAWLER FOR IOC IMPLEMENTED

FIGURE 9: TWITTER EVENT REPORTED IN THE S4R MISP

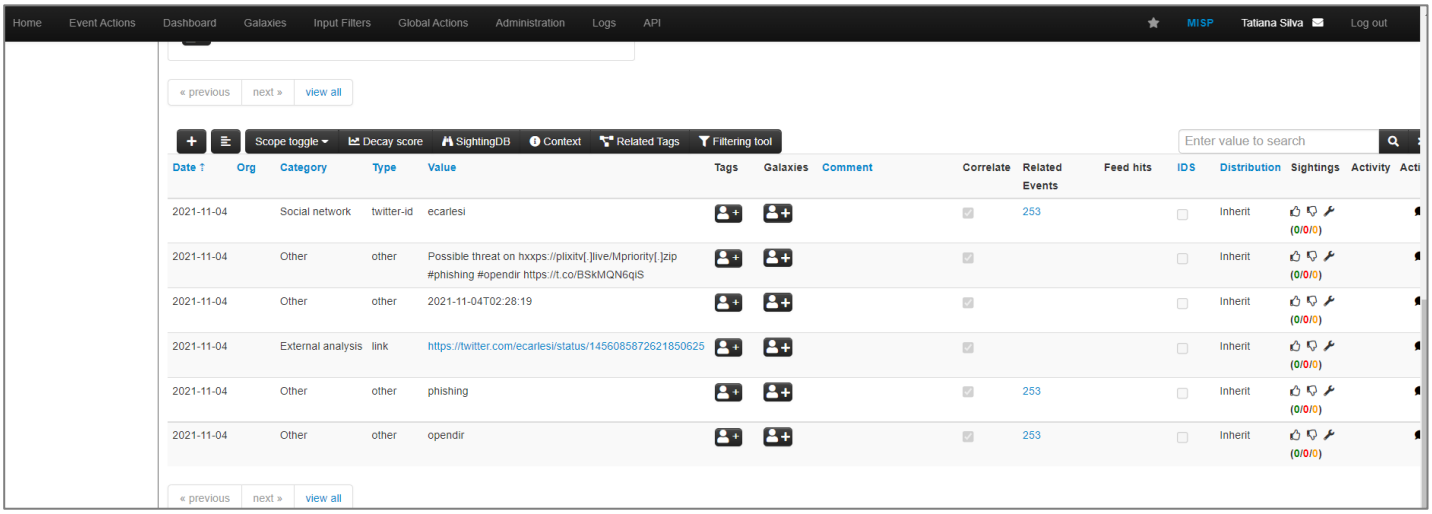


FIGURE 10: ALERT DETAILS AND LINK TO ORIGINAL TWITTER POST

DA-TIS-05 PHISHING CAMPAIGN MONITORING CRAWLERS

This tool allows tracking phishing campaigns using the tool dnstwist[17] in order to generate URL variants that are typically used in phishing campaigns (by mimicking typos etc., see Figure 11 for example base URLs and Figure 12 for example permutations generated wit dnstwist) and a simple database for storing the potential phishing campaigns in a MISP instance. The tool uses 2 main software components:

1) a Bash file that reads domain names and checks with dnstwist the existence of potential phishing campaigns supplanting the domain names of the list. The output is stored in JSON format in the report folder and the domains are stored in the database for traceability;

2) a Python script for parsing JSON reports by dnstwist and sending those domains that have not yet been saved into the database to in the database.

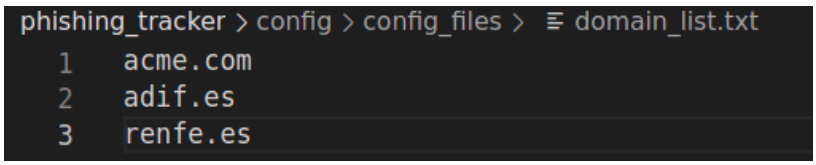


FIGURE 11: EXAMPLE DOMAIN NAMES MONITORED BY TISAIL

[17] dnstwist, https://github.com/elceef/dnstwist (accessed 31.03.2022)

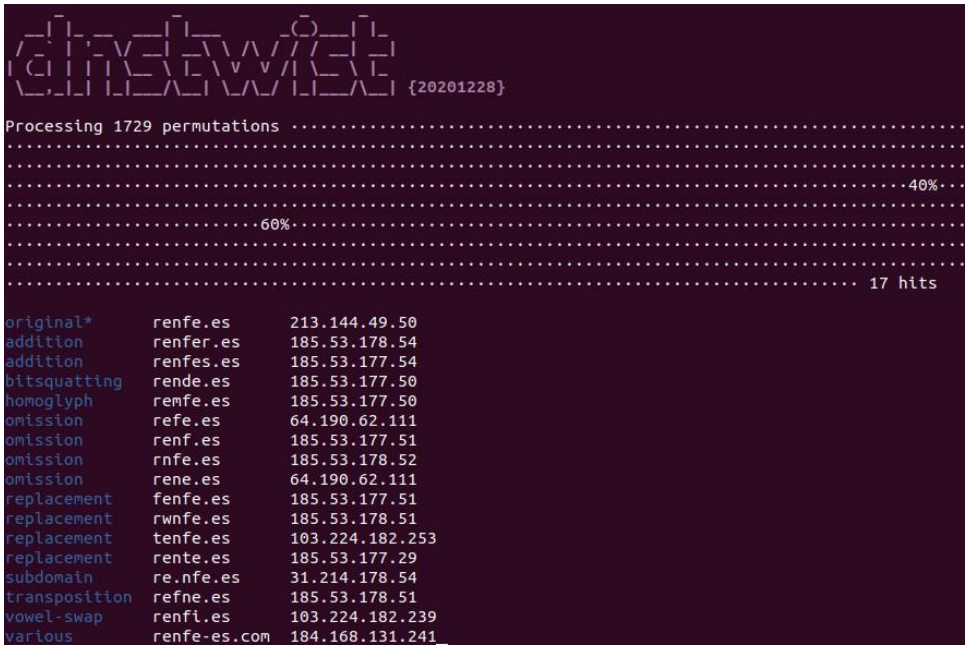


FIGURE 12: DOMAIN NAMES SIMILAR TO RENFE.ES

Potential attacks can then be stored in MISP for representation and propagation to tools that use the identified threat data (see Figure 12 and Figure 13).

FIGURE 13: DOMAIN NAMES MONITORED BY TISAIL

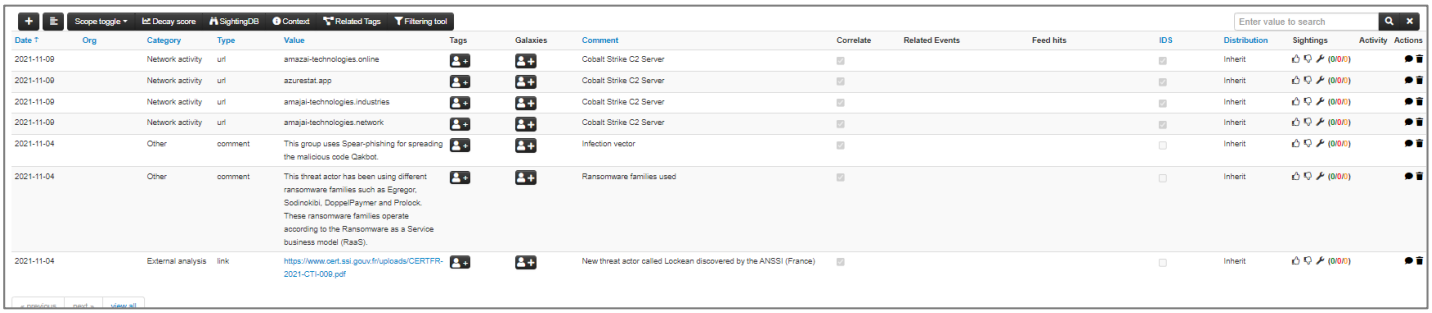


FIGURE 14: DETAILS FROM SPEAR-PHISHING CAMPAIGN

DA-TIS-06 THREAT INTEL FEED SELECTION AND CONFIGURATION

TISAIL uses threat intelligence feeds for situational threat awareness. The module created for this purpose allows gathering Indicator of Compromise (IoCs) from OTX Threat Exchange (see Section 4.2.1) subscriptions and sending them to MISP for threat analysis. The script uses the OTX Python SDK[18] to interact with OTX. The OTX platform contains many alerts provided by security researchers and is a convenient source for monitoring emerging cyber threats (see Figure 15 for examples).

[18] OTX Python SDK, https://github.com/AlienVault-OTX/OTX-Python-SDK (accessed 31.03.2022)

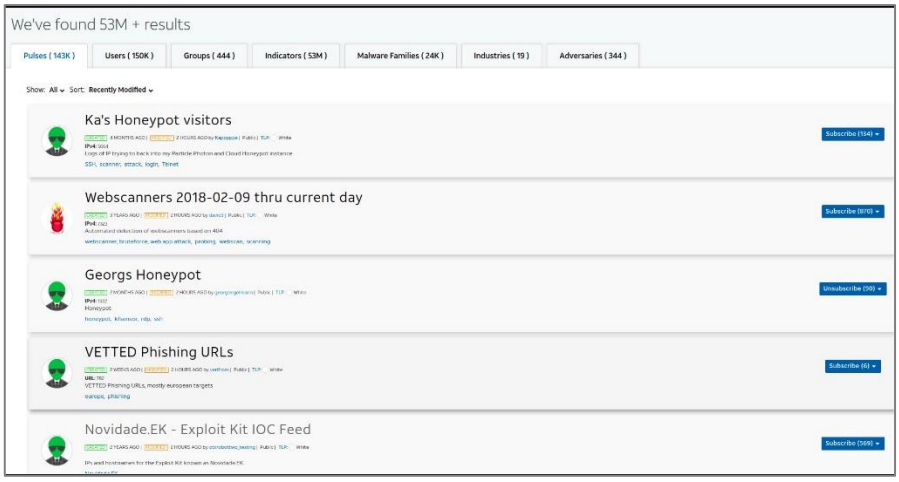


FIGURE 15: OTX DATA PLATFORM

The script gathers OTX threat updates and prints them to the console. By default, the script does not filter any retrieved threats, but the script allows us to narrow the number of threat categories gathered in 2 different ways:

- By keyword using -a parameter (e.g. Dridex, Web Shell, etc)
- By ATT&CK technique using -t parameter (e.g. T1078)

Figure 16 depicts the raw output generated by the OTX data gathering script and printed to the console.

```
[*] Searching for Pulses on OTX:
        [+] New OTX Pulse by AlienVault created: 2021-06-03T11:24:40.517000
                [-] This pulse was edited: 2021-06-03T11:24:40.517000
                [-] Title: SharpPanda Targets Southeast Asian Government With Previously Unknown B
                [-] ID: 60b8bbf9744408d3a83062f7
                [-] TLP: white
                [-] Description: Check Point Research identified an ongoing surveillance operation
                [-] Malware families: ['MClient', 'VictoryDll', 'RoyalRoad', 'Cobalt Strike - S015
                [-] Targeted countries: Unknown
                [-] Adversary: SharpPanda
                [-] ATT&CK Techniques: ['T1566.001', 'T1204.002', 'T1203', 'T1059.003', 'T1053', '
                [-] Tags: #cobaltstrike #SharpPanda #RoyalRoad #VictoryDLL #MClient #SharpM
                [-] References:
                        [+] link: https://research.checkpoint.com/2021/chinese-apt-group-targets-s
                [-] IoCs associated with this pulse:
                        [-] New IoC with ID: 3023387166
                                [-] IoC: d843b58f31c687d22de09a6765b3ba3b
                                [-] type: FileHash-MD5
                                [-] created: 2021-06-03T11:24:42
                                [-] title: ALF:Trojan:Win32/MeterLoad
                [-] New IoC with ID: 3023387180
                                [-] IoC: b40476638c83b8800413cf1fe88e28c2486367b79d1ddae7eb1ddcfa7
                                [-] type: FileHash-SHA256
                                [-] created: 2021-06-03T11:24:42
                                [-] title: Trojan:Win32/Tiggre!rfn
                                [-] description: SHA256 of 7a38ae6df845def6f28a4826290f1726772b247
```

FIGURE 16: GATHERING OTX PULSES

The alerts gathered via the OTX threat update script are sent to the MISP instance.

MISP Data Acquisition Implementation

The data acquisition components implemented for use with the SAFETY4RAILS MISP instance focus on the acquisition of data that is not covered by TISAIL, which focuses of the cyber threat domain but does not address OSINT data acquisition of threats from the physical domain. In order to support the acquisition of data from the physical domain, two data acquisition modules were developed in order to acquire data from structured data feed format data sources and social media data sources respectively:

- An RSS data feed acquisition module was developed in order to be able to acquire generic structured data feed information for import into the MISP database (DA-MSP-02). RSS format data feed messages can be parsed into custom data structures depending on the source feed URL.
- A Twitter[19] search data acquisition module was developed in order to be able to acquire data both from structured or semi-structured and unstructured social media data sources (DA-MSP-03). The data acquisition module can acquire data using searches via combinations of hashtags and keywords using the official Twitter API[20].

Import modules require the definition of target RSS feeds or social media feed sources or query terms in order to define the relevant data sources. Figure 17 shows an example query field for specifying a custom query for a new search query:

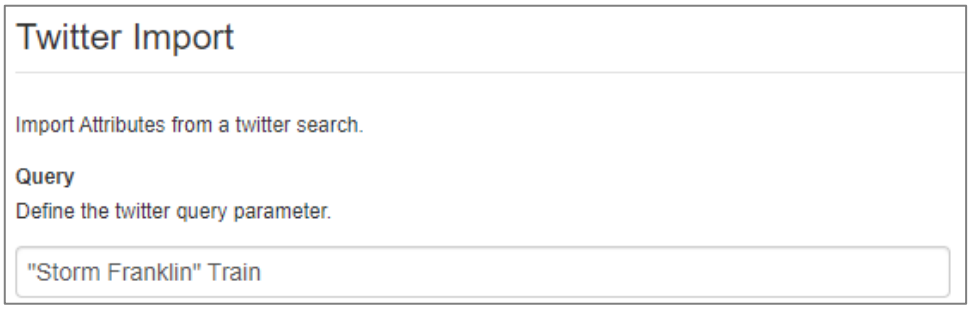


FIGURE 17: TWITTER QUERY DEFINITION EXAMPLE

Data that is retrieved via RSS and social media sources is analysed using a natural language processing (NLP) pipeline (described in D4.1) that extracts relevant information from the data that has been retrieved. The extracted data is then used to create MISP format events that are imported into the MISP instance connected to the module:

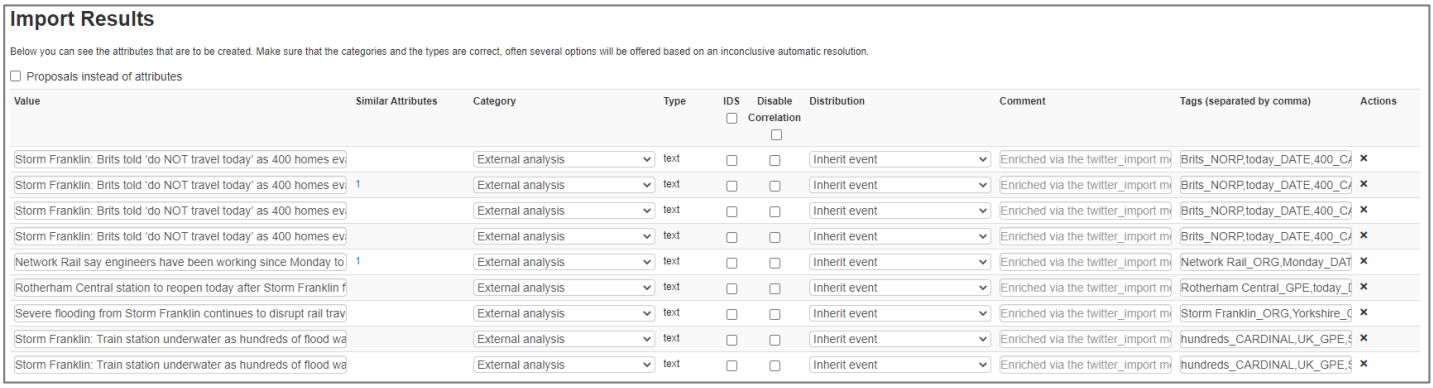


FIGURE 18: PARSED AND MAPPED IMPORT RESULT

[19] https://twitter.com/home (accessed 29.03.2022)

[20] Usage of the Twitter API requires successful completion of a formal approval process with Twitter, which Innova Integra successfully applied for within the scope of using the Twitter API during SAFETY4RAILS project activities.

The events that have been created can be viewed in the MISP user interface and can also be edited manually if that is desirable in a use case.

Date	Org	Category	Type	Value	Tags	Galaxies	Comment
2022-03-01		External analysis	text	Severe flooding from Storm Franklin continues to disrupt rail travel in Yorkshire https://t.co/ATAUzzcG7E #uk #railway #train https://t.co/AOqCiFaQ0e	Yorkshire_GPE, Storm Franklin_PERSON		Enriched via the twitter_import module
2022-03-01		External analysis	text	Rotherham Central station to reopen today after Storm Franklin flood https://t.co/XvW0nMdp4r #railway #rail #train	Rotherham Central_GPE, today_DATE, Storm Franklin_PERSON		Enriched via the twitter_import module
2022-03-01		External analysis	text	Storm Franklin: Train station underwater as hundreds of flood warnings in place across UK.A South Yorkshire rail station has been submerged under water as torrential rains from Storm Franklin cause travel chaos across the country.#Ukraine#يوم_القاسيس https://t.co/W1dhDLkxR9	hundreds_CARDINAL, UK_GPE, South Yorkshire_GPE		Enriched via the twitter_import module
2022-03-01		External analysis	text	Storm Franklin: Brits told 'do NOT travel today' as 400 homes evacuated & train stations underwater https://t.co/bIOZJ23HIN	Brits_NORP, today_DATE, 400_CARDINAL		Enriched via the twitter_import module
2022-03-01		External analysis	text	Network Rail say engineers have been working since Monday to pump water away from the railway between Aldwarke and Tinsley.https://t.co/tgo2L3kib9	Network Rail_ORG, Monday_DATE, Aldwarke_PERSON, Tinsley_PERSON		Enriched via the twitter_import module

FIGURE 19: RESULT EVENT ATTRIBUTES INCLUDING IDENTIFIED TAGS

The system attempts to extract a sufficient amount of data from the available input data using both NLP processing, available message metadata and supports custom template matching for semi-structured data sources such as specifically formatted RSS feed sources. The figure below again illustrates how data from a social media post is mapped into a MISP event object.

Twitter-post Object Template

ID	257
Name	twitter-post
Organisation	ORGNAME
UUID	d1214031-ce1b-4a35-bd33-644c707bda2e
Version	5
Meta-category	misc
Description	Twitter post (tweet).
Requirements	**requiredOneOf** post post-id archive url link attachment

FIGURE 20: TWITTER-POST MAPPING INCLUDING TAGS

Figure 21 illustrates the analogous data mapping for an RSS feed event generated from a semi-structured RSS feed data source that indicates status and changes to national threat levels as published by the UK MI5 security service.

RSS Feed

```
<rss xmlns:atom="http://www.w3.org/2005/Atom"
     version="2.0"
     xml:base="https://www.mi5.gov.uk/">
 <channel>
  <title>Threat Level</title>
  <link>https://www.mi5.gov.uk/</link>
  <description>The Current UK Threat Level</description>
  <language>en-gb</language>
  <copyright>© Crown Copyright</copyright>
  <generator>MI5</generator>
  <lastBuildDate>Wed, 09 Feb 2022 14:33:49 +0000</lastBuildDate>
  <ttl>20</ttl>
  <item>
   <title>Current Threat Level: SUBSTANTIAL</title>
   <link>https://www.mi5.gov.uk/threat-levels</link>
   <description>
      The current national threat level is SUBSTANTIAL. The
      threat to Northern Ireland from Northern Ireland-related
      terrorism is SEVERE.
   </description>
   <pubDate>Wednesday, February 9, 2022 - 14:33</pubDate>
   <subject>ThreatLevel</subject>
  </item>
 </channel>
</rss>
```

MISP RSS-item Object

```
{
        "name": "rss-item",
        "meta-category": "misc",
        "template_uuid": "3eaaadd2-429e-4202-a669-b1b98817e928",
        "description": "RSS Item",
        "template_version": "20220228",
        "uuid": "623ef7d2-a432-4608-ba96-79186907abf8",
        "Attribute": [{
                "uuid": "bf843d97-f10f-479c-8bbd-80e961272287",
                "object_relation": "title",
                "value": "Current Threat Level: SUBSTANTIAL",
                "type": "text",
                "disable_correlation": true,
                "to_ids": false,
                "category": "Other"
        }, {
                "uuid": "92205882-4432-426d-bd2c-3cd69f02f3d1",
                "object_relation": "link",
                "value": "https://www.mi5.gov.uk/threat-levels",
                "type": "link",
                "disable_correlation": false,
                "to_ids": false,
                "category": "External analysis"
        }, {
                "uuid": "6aaf340b-0a99-46e3-ab11-41ce46258948",
                "Tag": [{
                        "name": "Northern Ireland_GPE"
                }],
                "object_relation": "description",
                "value": "The current national threat level is SUBSTANTIAL. The
                          threat to Northern Ireland from Northern Ireland-
                          related terrorism is SEVERE.",
                "type": "text",
                "disable_correlation": true,
                "to_ids": false,
                "category": "Other"
        }, {
                "uuid": "5762260e-5ccf-4276-a097-3ee2497e2be0",
                "object_relation": "pubDate",
                "value": "2022-02-09T14:33:00",
                "type": "datetime",
                "disable_correlation": true,
                "to_ids": false,
                "category": "Other"
        }],
        "distribution": "5",
        "sharing_group_id": "0"
}
```

FIGURE 21: RSS FEED MESSAGE TO MISP EVENT MAPPING

Both the RSS and Twitter import modules were developed in Python as custom MISP import modules which can be added to MISP installations via the MISP module extension mechanism[21] and can then be operated in the same manner as built-in MISP functionalities in order to retrieve data in regular intervals.

3.2.2 Pre-Processing and Analytics

Data Pre-processing and analytics components handle incoming data and process in order to select, convert, verify or otherwise evaluate or convert the original data into data that are suitable for input into the system database and relevant for use in the application domain. Table 5 lists the pre-processing and analytics components.

TABLE 5: PRE-PROCESSING AND ANALYTICS COMPONENTS

ID	*Title*	*Description*
PA-MSP-01	**Threat intel feed automated relevance filtering**	Development of a filter rule system to filter out threat intelligence that does not relate to assets/components that has been declared as being in use by railway domain partners
PA-MSP-02	**Structured physical threat automated data feed filtering**	Development of a filter rule system to filter out structured threat data that does not relate to the relevant assets declared by railway domain partners
PA-MSP-03	**Semantic social media data parsing for physical threat detection**	Extraction of entities from social media messages in order to identify messages that may relate to the relevant assets declared by railway domain partners

Pre-processing of threat intel and physical threat data is achieved through

- custom specifications of relevant systems and devices for the domain (using example data provided by transport user partners in the project),
- specification of representative example domains relevant for domain spoofing and similar domain-specific custom attacks and
- selection of data according to geographic search parameters for both structured and unstructured data sources in order to identify data relevant to a specific geographic area where relevant.

Both TISAIL and the SAFETY4RAILS MISP implementation use configurations that specify relevant search properties using these parameters, which are integrated into the configuration of data acquisition components described above.

[21] https://misp.github.io/misp-modules/contribute/ (accessed 29.03.2022)

3.2.3 Storage and Representation

Storage and data representation components manage and model the data gathered and processed in data gathering and pre-processing steps. Table 6 lists the storage and data representation components.

TABLE 6: STORAGE AND REPRESENTATION COMPONENTS

ID	*Title*	*Description*
SR-TIS-01 (TISAIL)	**Data repository operation**	Secure operation of the TISAIL data repository for data persistence and management
SR-TIS-02 (TISAIL)	**Data model customisation**	Customisation of the TISAIL data model in order to suitably represent relevant threats (based on existing data model used with TISAIL)
SR-MSP-01	**Data repository operation**	Secure operation of the SAFETY4RAILS MISP repository for data persistence and management
SR-MSP-02	**Data model customisation**	Development of an integrated data model for cyber, physical and cyber-physical domain elements included assets/components modelled as MISP objects with taxonomy linkage

Identifiers for data acquisition components contain “TIS” when they are implemented as part of TISAIL in order to differentiate them from components that are not implemented directly as part of that system. Identifiers for components contain “MSP” when they are implemented as part of the project MISP instance. SR-TIS-02 and SR-MSP-02 are described in Section 4 as part of the overall data model description.

A TISAIL instance as well as a SAFETY4RAILS MISP instance were deployed and hosted for use in the SAFETY4RAILS project as described in Section 3.3.2.

3.2.4 Data Set Analytics

Data set analytics components compute analytics over the overall dataset available in the open source intelligence database. This includes processing when specific new or new types of data points are added and processing in specified intervals in order to e.g. generate overall statistics or ranking data. Table 7 lists the data set analytics components.

TABLE 7: DATA SET ANALYTICS COMPONENTS

ID	*Title*	*Description*
DS-MSP-01	**Rule evaluation over database triggered by specific conditions**	Integration of a rule engine or similar mechanism that enables the evaluation of rules when specific changes to the database have been detected
DS-MSP-02	**Generation of summary statistics**	Development of a component that generates summary statistics for threats and vulnerabilities including the generation of ranked lists of threats and vulnerabilities for a given time period

DS-MSP-01 is implemented by deploying and using Yara (see Section 3.2.1) rule processing both for malware and rule-based event data and metadata processing. MISP provides both built in and advanced customisable support for Yara rules, including options to conditionally export data from MISP for processing with Yara rules. TISAIL uses Yara rules to analyse incoming threat messages.

DS-MSP-02 is implemented as part of a custom Java microservice that monitors all threats and vulnerabilities that are reported as relevant through the S4RIS DMS.

Summary statistics concerning the most frequently reported types of MISP events (using tagging data to identify types of threats identified) are generated in regular intervals and are summarised on a daily, weekly, monthly basis as well as for the overall runtime of the service. From the software point of view, the function is implemented by a hash map counting the occurrences of threat classes and vulnerability classes.

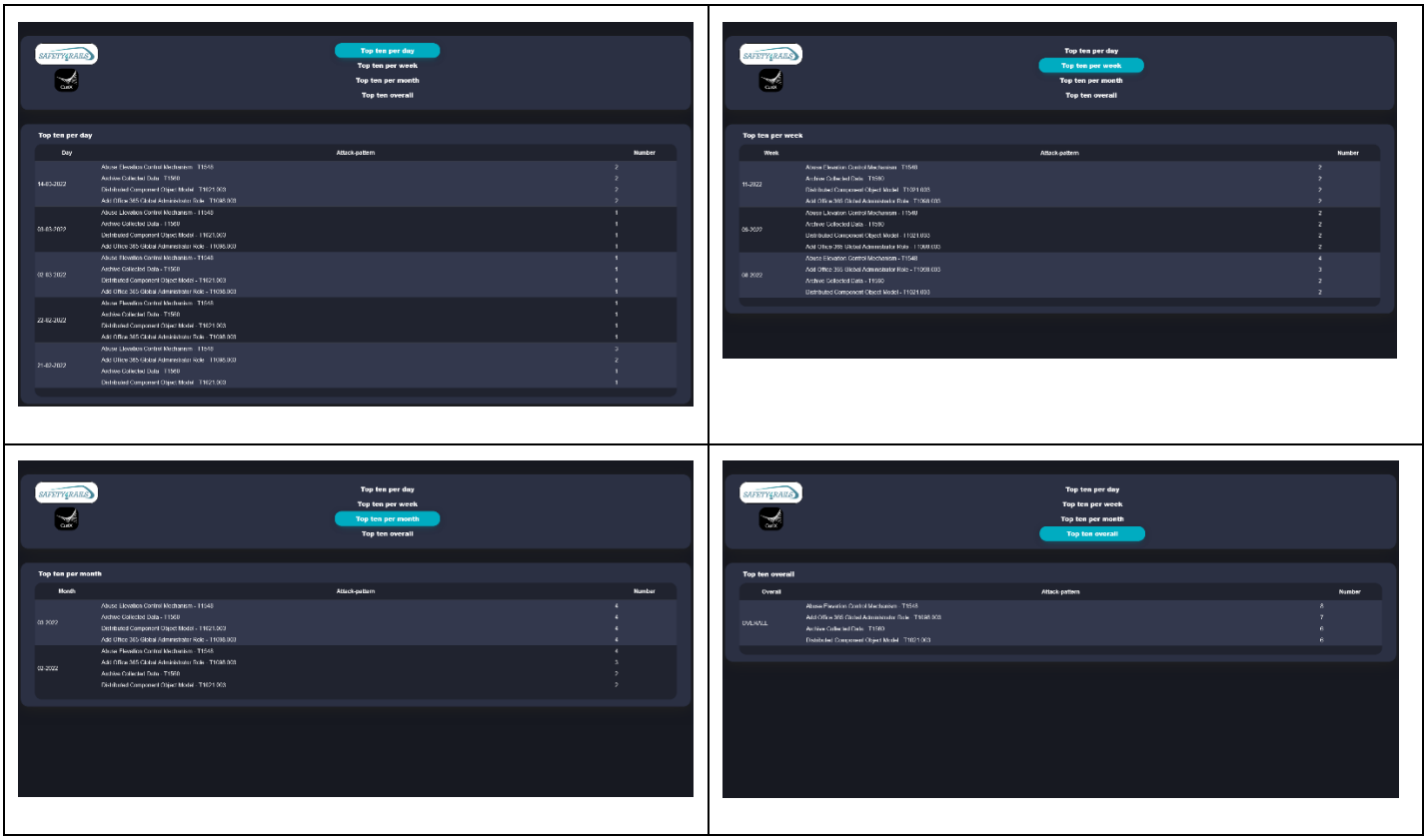


FIGURE 22: DAILY, WEEKLY, MONTHLY AND OVERALL LISTS OF MOST FREQUENTLY ENCOUNTERED THREATS

A REST API endpoint can be polled to retrieve a sorted list of those threat classes and vulnerability classes. A simple web user interface has been developed in order to present the generated ranked lists (see Figure 22).

3.2.5 Data Access and Messaging

Data access and messaging components enable access to the database containing gathered and processed OSINT data either upon request or proactively through the message broker when so specified. Table 8 lists the currently specified data access and messaging components.

TABLE 8: DATA ACCESS AND MESSAGING COMPONENTS

ID	*Title*	*Description*
DM-TIS-01	**Synchronisation mechanism with SAFET4RAILS MISP instance**	Integration of a synchronisation mechanism that synchronises data generated by TISAIL into the SAFETY4RAILS MISP instance
DM-MSP-01	**Custom API endpoint development for**	Development of a component that operates an extended MISP API endpoint with convenience methods that reduce the complexity of requests for frequent complex API client requests

	typical complex queries	
DM-MSP-02	**Message broker integration for SAFETY4RAILS system communication**	Integration and configuration of a client that integrates MISP with the project-wide message broker system

Identifiers for data acquisition components contain “TIS” when they are implemented as part of TISAIL in order to differentiate them from components that are not implemented directly as part of that system. Identifiers for components contain “MSP” when they are implemented as part of the project MISP instance.

TISAIL-MISP synchronisation is achieved via a standard MISP server to MISP server synchronisation mechanism available via the TISAL backend.

The component envisioned in DM-MSP-01 was found to not be necessary, as the existing MISP API was found to be sufficient for use within the task and external communications are handled using the SAFETY4RAILS Apache Kafka-based messaging system. It was thus not necessary to implement any convenience methods for the API for use in SAFETY4RAILS.

A dedicated connector software module has been developed in order to connect the OSINT subsystem to the SAFETY4RAILS DMS. The connector software module is implemented in Java and consists of two main parts: Within SAFETY4RAILS, the software tool CuriX (by IC) should be enabled to address Real time monitoring of cyber threats (see D1.4, Requirement CuriX_03). CuriX already has a proprietary library which allows the connection to several REST based APIs. This library was extended with a connection mechanism to the MISP instance. Technically, GET requests can be sent to a MISP API and JSON responses can be consumed by using this library. Furthermore, this library was extended to allow sending events to the SAEFTY4RAILS DMS using the DMS’s RESTful API.

A new Java microservice was developed, using the aforementioned library to orchestrate requests to and responses from MISP. Based on a configurable polling mechanism, GET requests are sent regularly to the MISP instance. Responses from MISP are then transformed into a JSON format compatible with the JSON message definitions used for messaging via the DMS and sent via REST to a configurable DMS topic. Figure 23 shows a top-level view of the implemented service integrated as interfacing service between MISP and the DMS.

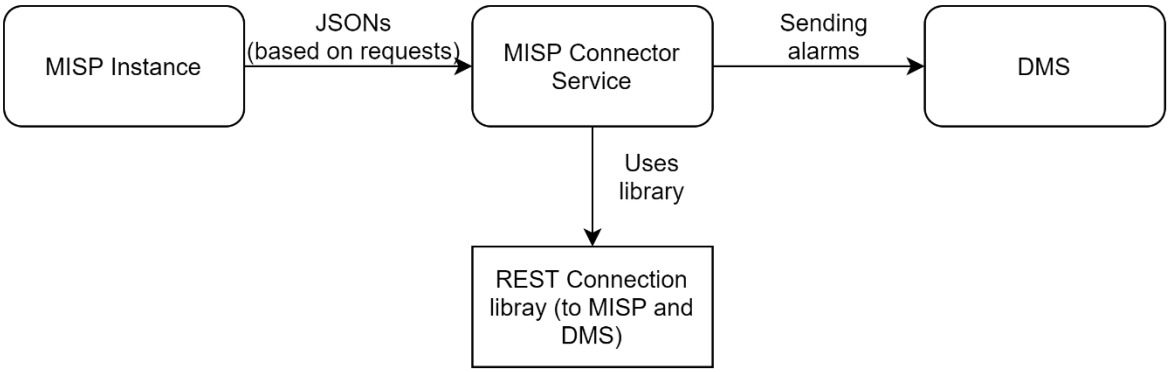


FIGURE 23: IMPLEMENTED SERVICE THAT TAKES VULNERABILITIES AND THREATS FROM MISP TO DMS

3.3 Architecture and Deployment

This subsection describes the architecture implementation used for the OSINT subsystem and the relevant constituent components as well as the component deployment that was used during the project for use in the various simulation exercises of the SAFETY4RAILS project.

3.3.1 Architecture

The system architecture of the OSINT system has been derived from the processing pipeline structure and the division of system components into TISAIL and generic MISP components. Figure 24 presents a conceptual component diagram of the overall OSINT system. The architecture, diagram and this initial subsubsection remain largely unchanged relative to D4.3 since the architecture specification was not changed during implementation apart from the removal of a custom API for convenience methods (see below).

The diagram shows a conceptual view where components are labelled with the component IDs given in the component description tables in Section 3.2. Multi-component diagram elements are labelled with non-specific identifiers such as “DA-TIS-XX” in order to denote that all of the types that belong to the family of components denoted by the remainder of the ID (e.g. DA-TIS-01, DA-TIS-02, …) are shown there as a group. Components that are not software components are added to the diagram as notes (these are data models and the message integration of the TISAIL repository and the MISP repository). The “open source world of available data” is indicated using the cloud symbol outside of the overall OSINT system border.

Please note that the TISAIL equivalents of the “PA-MSP-XX” and “DS-MSP-XX” components for the TISAIL system are also part of the overall OSINT system but are already integrated as part of the TISAIL repository and are hence not listed here as separately developed or integrated components.

The architecture integrates the proprietary TREE TISAIL solution into the overall system architecture in a simple and practical way that enables TREE to securely host, use and further develop their system as part of the project work and that allows the other technical partners in the project to develop and contribute their own IP and developments to the project through the project MISP repository. Additional overhead caused by this remains minimal by using MISP instance synchronisation between the TISAIL repository and the project MISP instance via a built-in MISP synchronisation feature.

The Broker Client extends outside of the boundary of the OSINT system in order to indicate that the system integrates with the designated message broker system that is used in the SAFETY4RAILS system. As noted earlier, the originally architected API façade implementation for the OSINT (DM-MSP-01) system has not been implemented as all communications with external systems are handled through the message broker client. The revised Figure 24 has been updated to reflect this change accordingly.

The consumers of the OSINT data processing activities carried out in the system within SAFETY4RAILS are primarily Task T4.5 concerned with the implementation of real-time monitoring, Task 5.5 concerned with the integration of RAM^2, and the overall S4RIS information system that is managed in WP6. More detailed specifications regarding this can be found in deliverable D1.4 “Specification of the Overall Technical Architecture” (please note that this is a document with limited public access). Finally, it should be noted that while specific consuming systems are part of the SAFETY4RAILS project and their needs have been considered during the development of the OSINT system, the system remains open and accessible by all systems that can be integrated into the overall SAFETY4RAILS infrastructure.

We close with some remarks concerning the practical implementation of the system in terms of use of programming languages and similar matters. The relevant boundary conditions for the technical implementation of the OSINT system are primarily determined by the interfaces and access functionalities provided by the MISP system. MISP provides both a REST API and an API access wrapper library written in Python. In addition, MISP uses standard backend systems including a MySQL database backend to which developers can gain direct access in order to carry out modifications that are not supported by the MISP system itself.

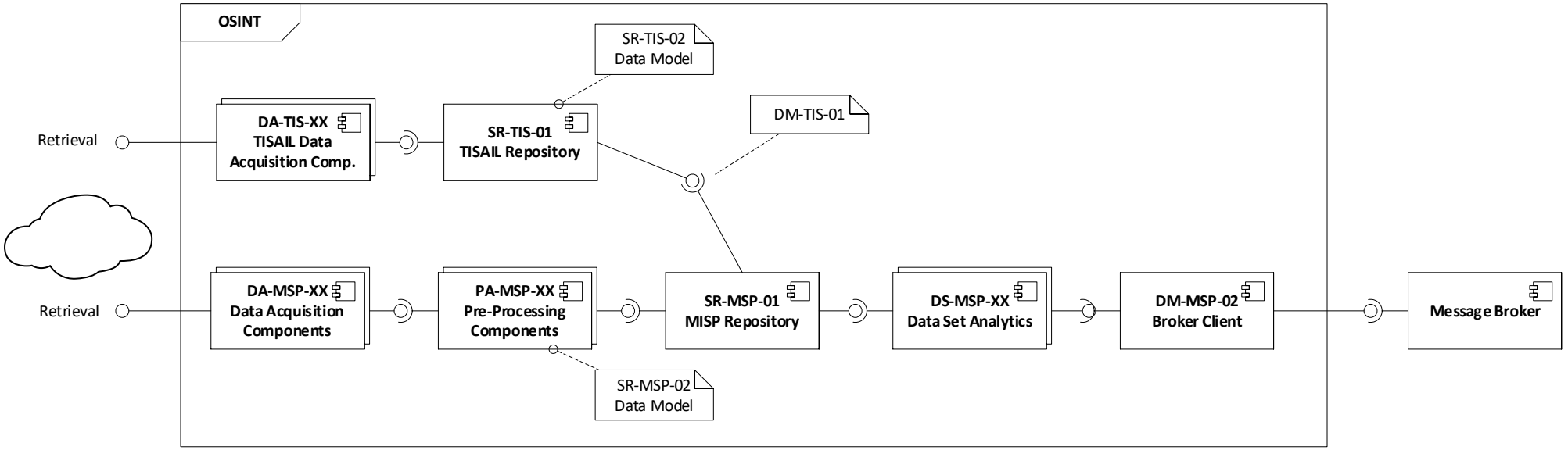


FIGURE 24: OSINT SYSTEM ARCHITECTURE COMPONENT DIAGRAM

3.3.2 Deployment and Hosting

The OSINT system was deployed for being used in testing, validation and demonstration activities as part of SAFETY4RAILS. Two main deployments were hosted separately: a TISAIL instance configured for use within the project, and a dedicated MISP instance to gather and propagate data from both TISAIL and other data sources.

TISAIL Instance Deployment and Hosting

TISAIL has been deployed to a protected server environment hosted at Tree Technologies. External connectivity both for the purpose of data acquisition and communication to the T4.2 MISP instance was restricted using typical security mechanisms in order to ensure secure deployment of the system.

TISAIL uses MISP which provides different alternatives for deploying the platform such as Docker or cloud-ready images for deploying MISP on cloud providers. TISAIL has been deployed using a cloud-ready image for AWS provided via the official MISP repository. The image can be found on the AWS AMI repository (see Figure 25).

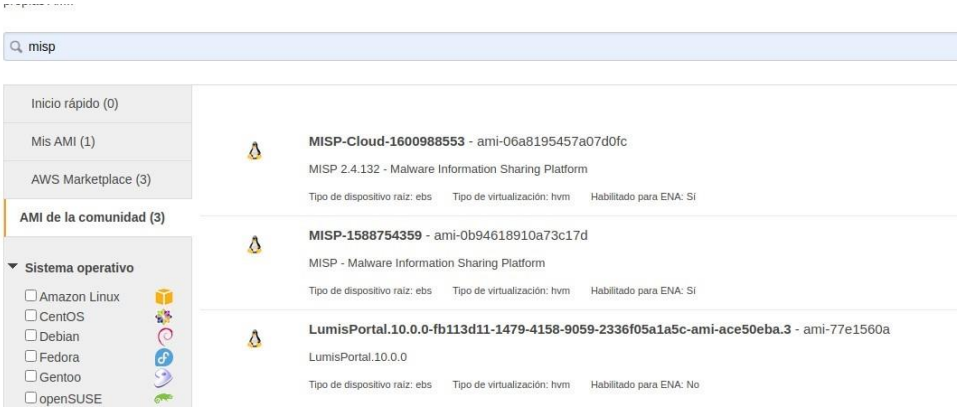


FIGURE 25: AWS AMIS

The MISP instance is hosted on a private subnetwork in order to avoid for the instance to be reached from the open Internet. The figure below shows the architecture of the Threat Intelligence infrastructure.

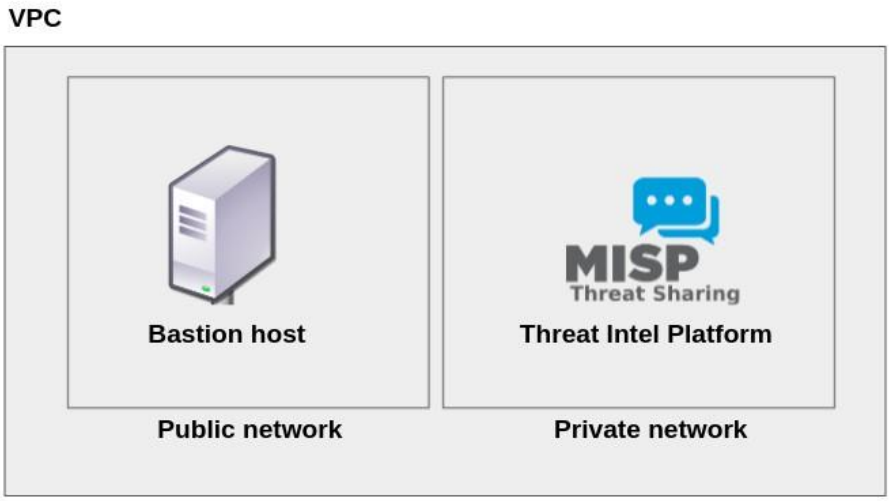


FIGURE 26: TISAIL ARCHITECTURE

TISAIL has been deployed in a Virtual Private Cloud (VPC) within Amazon Web Services (AWS). The TISAIL platform is hosted in an instance in a private subnetwork, thus avoiding being exposed to the Internet and only being reachable by authenticated users. Access is mediated using a specifically hardened “bastion” host in the public subnet that controls access to the private subnet. The bastion host only accepts connections from known IP addresses and public-key authentication is required for access.

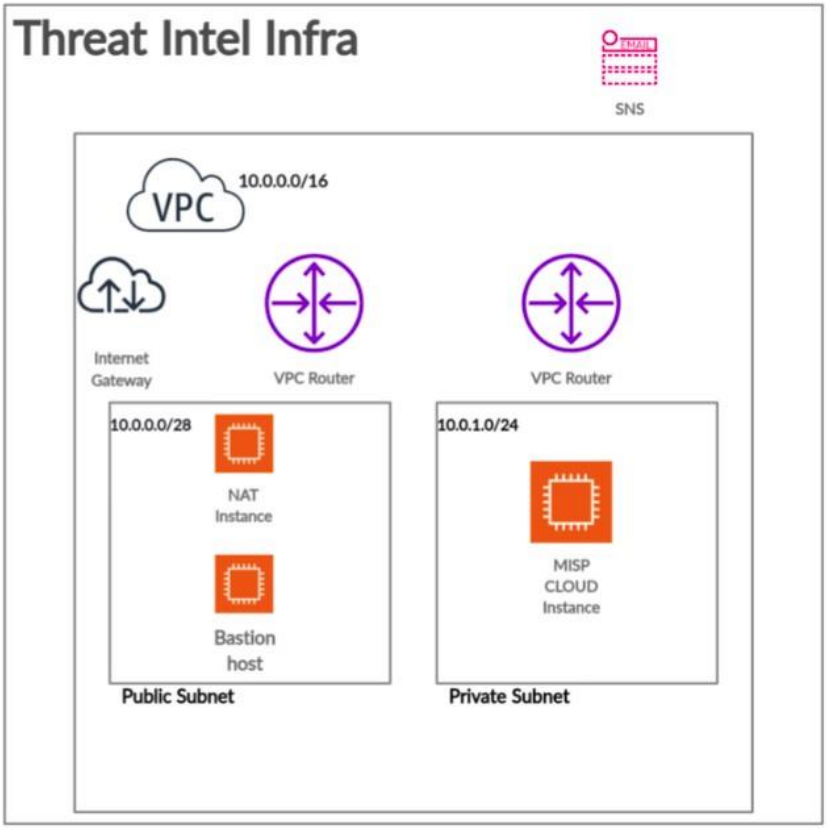


FIGURE 27: TISAIL AWS ARCHITECTURE

The bastion host-access security group contains the list of IP addresses that are allowed to connect to the Bastion-host via SSH. Using an SSH-tunnel, the user can access the MISP instance via SSH for administrative tasks or via web using a proxy SOCKS5.

Project MISP Instance Deployment and Hosting

The project MISP instance has been deployed on a virtual machine provided by and also regularly updated by CyberServices. The MISP virtual machine instance runs on Ubuntu Server LTS 20.04 operating system. The MISP instance is accessible by partners over the web interface and through the default MISP API with credentials provided to the task participants that require access.

The version of the deployed MISP instance is 2.4.150. The main configuration changes to a default installation are for user access and deployment of custom taxonomies. User access was provided to Innova Integra, Tree Technology and CuriX for their respective integration activities. A taxonomy provided by Tree Technology has been installed on the MISP backend. The taxonomy is used in the synchronisation process between the MISP and the TISAL instances.

Partner Module Deployment and Hosting

Partners Innova Integra and CuriX have deployed custom modules for data acquisition and data messaging with the S4RIS Platform respectively. These modules are hosted independently of the MISP instance deployed for T4.2. They connect to the instance via the available MISP API and are secured via the authentication and authorisation mechanisms provided by MISP and the S4RIS Platform distributed messaging system implementation respectively.

The Innova Integra data acquisition components have been deployed to a separate server instance hosted by Innova Integra, which have been integrated with the T4.2 MISP instance via the MISP remote plugin configuration mechanism. Communication is secured using PKI security supported by MISP and exchanged between Innova Integra and CyberServices.

4. Data Model, Data Sources & Data Acquisition

This section introduces the abstract data model that defines the key entity types that are relevant in the context of SAFETY4RAILS OSINT. Furthermore, this section lists basic data sources and data acquisition requirements that need to be met within the project in order to enable the task to successfully carry out the defined activities.

Many of the core elements described in this section remain unchanged relative to D4.3 and have been carried over from that deliverable. We have integrated sections describing the models implemented for use in MISP and elsewhere as appropriate.

4.1 Data Model

The "world of concern" for SAFETY4RAILS generally involves both cyber and physical assets and their security. The purpose of open source intelligence in the project is to improve the security of these assets by gathering information about potential vulnerabilities, past, present and future threats and to identify assets that have been compromised by a threat where that is possible using open source data.

4.1.1 Threat & Vulnerability Modelling Approach

Two specific challenging properties of this problem are

- the inclusion of cyber and physical security-relevant open source intelligence in a single system and
- covering as much as possible of the highly diverse and multi-faceted cyber and physical infrastructure that exists within railway operators.

Because of the vast and diverse array of possible assets and related security issues, we view the data to be gathered via OSINT from a high-level perspective at theoretical level. Figure 28 illustrates this simple overall model.

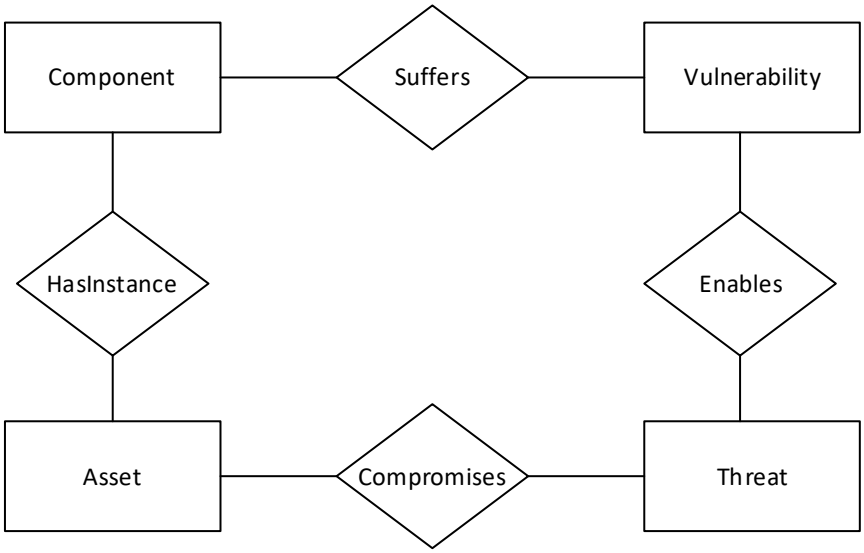


FIGURE 28: HIGH-LEVEL ER-DIAGRAM OF OSINT ENTITIES OF CONCERN

In this diagram, we reduce the view of our "world of concern" to the following four entities:

1. **Components** are specific and uniformly describable classes of "things" that are relevant within the context of railway security. Examples for a component are for instance a specific model of a SCADA component used within the railway infrastructure.
2. **Assets** are specific instances of components that are uniquely identifiable and are deployed in or potentially connected to the railway infrastructure of concern. Examples for an asset would be an actual SCADA system deployed within a railway infrastructure.

3. **Vulnerabilities** describe known risks of compromise or weaknesses that can be exploited by threats. Vulnerabilities can be defined at different levels: a direct connection to the Internet in itself could for instance be considered to be a vulnerability of a system.
4. **Threats** describe potential attacks or other types of threats to assets that may cause adverse effects to assets with relevant vulnerabilities.

This basic model is an example of a threat-driven data model (7) (8). It can be extended with representations of adversaries and effects of a threat (which can be referred to as an attack when it is executed) and related to observable events, attacks and/or incidents.

For the purpose of SAFETY4RAILS, data modelling activities have focused on profiling assets (and their generic super class of components) in terms of vulnerabilities and on identifying threats that may impact assets via component-level vulnerabilities, here specifically focusing on physical threats as the standard MISP data model already provides extensive coverage for cyber threats and vulnerabilities, which makes the MISP cyber threat and vulnerability base model the starting point for implementing an extended data model.

As part of this modelling, we have integrated the threat taxonomy that has been developed in T3.1 and was documented in D3.1 as the categorisation format for defining physical threat types. This allows for threat types to be mapped to the data retrieved from semi-structured data sources that are typically not explicitly marked up in terms of threat types (unlike in typical cyber threat report processing). The threat taxonomy created in T3.1 is also used in T3.4 to adapt SecuRail for SAFETY4RAILS, so that using the same mappings and identifiers enables interoperability and easier integration of the OSINT and SecuRail systems in the project.

4.1.2 MISP Data Modelling

MISP is primarily a system that receives and processes messages that provide information on (mostly cyber security) threats, incidents and elements connected to those in the form of events. (9) provide and describe a basic example of a MISP format event as follows:

```
{
    "Event": {
        "date": ":2002-03-12",
        "threat_level_id": "1",
        "info": "testevent",
        "published": false,
        "analysis": "0",
        "distribution": "0",
        "Attribute": [{
            "type": "domain",
            "category": "Network activity",
            "to_ids": false,
            "distribution": "0",
            "comment": "",
            "value": "test.com"
        }]
    }
}
```

FIGURE 29: EXAMPLE OF A MISP THREAT EVENT WITH A SINGLE ATTRIBUTE (9)

Beyond events, MISP data models can contain objects and can be structured and further defined using taxonomies, which are also used in order to tag events with relevant information. MISP galaxies[23] are more complex data model environments that can bundle and group larger data sets with complex relations that can

[23] See https://www.misp-project.org/galaxy.html (accessed 19.04.2021) for an overview over standard galaxies that are available with MISP.

be used dynamically in MISP (10). Galaxies and taxonomies can be used in order to model complex domains such as the one envisioned for the SAFETY4RAILS project.

While the MISP format is flexible and easy to extend, it also provides both a broad coverage of relevant event types, attributes and related data types and specifies and implementations of formalised concepts such as the estimative language model of likelihoods and confidence in sources, data and methodologies as defined in ICD 203 and JP 2-0 [24]. Use of such formalised concepts is also of key importance for the specification of concepts such as the impact of a threat or an executed attack on an asset.

4.1.3 Integration into SAFETY4RAILS Data Model Environment

SAFETY4RAILS has undertaken substantial work in the areas of defining and elaborating concepts for railway security both in the cyber and physical domains, including identifying critical IT & OT components in the railway domain (T3.1) as well as in particular threats (T3.2, T3.3). Risks and vulnerability modelling and assessment have been investigated in T5.1. All of these activities have both gathered representative instances and characterised them with attribute descriptions and taxonomies in order to organise and structure them into data models.

Since the timings of these activities and the development of the OSINT system overlapped, data model work in Task 4.2 initially focused on developing methods for the acquisition of open source data and the development and use of a lightweight taxonomy and model that can easily be reorganised to correspond to the final models produced by tasks T3.1, T3.2, T3.3 and T5.1. To this end, work in the task was initially carried out with modelling data that were created as part of the Shift2Rail EU initiative and there in particular the X2RAIL-1 project[25] and the 4SECURail project[26] and some of the MITRE ATT&CK® matrices[27].

In a second phase near the end of Task 4.2, output from WP3 and WP5 that was available was integrated into the description model of the OSINT system via MISP model extensions where applicable, specifically including threat data models from T3.1 and output format synchronisation for integration with the RAM^2 tool integrated as part of the activities of WP5.

4.1.4 MISP Model Usage and Extension

This subsection details the MISP model definitions that were created in order to integrate specific types of events that are not typically propagated via MISP to the default data model and in order to integrate the MISP data model with the SAFETY4RAILS data model environment.

For both social media and data feed data sources, integration with SAFETY4RAILS taxonomies is achieved by identifying and extracting terms from the provided data that is associated with taxonomy entries defined in the threat taxonomy documented in deliverable D3.1, focusing on identifying terms that are relevant for the simulation exercises in particular. The extracted terms are identified in the content that is retrieved using the approach described in deliverable D4.1 and associated with the data retrieved for analysis.

Twitter Data Representation

The representation of Twitter data in SAFETY4RAILS uses the MISP microblog object definition as described in the MISP documentation[28]. The example documented in Figure 17 depicts an example for Twitter data representation using this object definition data format.

---

[24] See https://www.misp-project.org/taxonomies.html#_estimative_language (accessed 19.04.2021) for the natural language description of this concept.
[25] See https://projects.shift2rail.org/s2r_ip2_n.aspx?p=X2RAIL-1 (accessed 19.04.2021) for further information on the X2RAIL-1 project.
[26] See https://www.4securail.eu/ (accessed 19.04.2021) for further information on the 4SECURail project.
[27] See https://attack.mitre.org/matrices/enterprise/ (accessed 25.05.2021) for further information on MITRE ATT&CK®.

[28] See https://www.misp-project.org/objects.html#_microblog (accessed 24.02.2022) for further information on the specification.

A custom object definition was defined for use with RSS format data acquisition. The object format identifies title, url, description, author, category, comments and sources as the main additional object properties in addition to default ones in order to represent RSS data objects in MISP.

4.2 Data Sources & Data Acquisition

In this subsection we briefly present the data sources for cyber security and physical security respectively and outline the framework for the development of the test and demonstration data for the task.

4.2.1 Open Source Cyber Security Data

A number of key services provide cyber security information that is relevant for the railway domain. The majority of these services provide access to general cyber security data, which need to be filtered in order to identify the subset that is relevant for the railway domain. Note that this subset is still likely to include a range of general cyber security threats, for instance ones relating to information displays and terminals operating legacy Microsoft Windows operating systems.

TABLE 9: OPEN SOURCE CYBER SECURITY DATA SOURCES

Name	*Description*	*References*[29]
VirusTotal	Malware repository	https://www.virustotal.com
PolySwarm	Malware and online threat detection repository	https://polyswarm.io
MalwareBazaar	Malware repository	https://bazaar.abuse.ch
Hybrid Analysis	Malware repository	https://www.hybrid-analysis.com
Any.run	Malware analysis sandbox and repository	https://any.run
Shodan	Network security search engine	https://www.shodan.io
Censys	Internet-wide scanning service	https://censys.io
ZoomEye	Address and port scanning service	https://www.zoomeye.org
OTX Threat Feed	Threat Intel feed by AT&T (Former AlienVault)	https://otx.alienvault.com
DigitalSide Threat Intel	Regular data feeds on threat intelligence for cyber security	https://osint.digitalside.it
CERT data feeds	Regular data feeds on breaches, vulnerabilities and system misuse	List of EU CERTS at https://www.enisa.europa.eu/

[29] All web resources accessed 27.03.2022.

Twitter	Public feeds of threat alert accounts that are operated by organisations and that are not individual user accounts	https://twitter.com
CISA National Cyber Awareness System	Cyber threat bulletin services by the U.S. government	https://www.cisa.gov/uscert/ncas

In addition to data repositories and search engines, additional relevant domain data for the identification of frequently used railway domain security-relevant websites and relevant security-related social media accounts have been specified for the demonstration of phishing specialising in the railway security domain.

4.2.2 Open Source Physical Security Data

Open source physical security data is generally not as readily available via structured online data services and data sources as open source cyber security data.

TABLE 10: OPEN SOURCE PHYSICAL SECURITY DATA SOURCES

Name	*Description*	*References*[30]
GTD Global Terrorism Database	Regularly updated database of terrorist incidents	https://www.start.umd.edu/gtd/
Twitter	Public feeds of threat alert accounts that are operated by organisations and that are not individual user accounts	https://twitter.com
News agency news feeds	News agency news feeds for local events including weather and traffic events	https://www.reuters.com
European weather warning feed	Feed with weather warning data for meteorological events	https://meteoalarm.org
MI5 terrorism threat level information feed	Feed notifying with changes to the domestic terrorism threat level	https://www.mi5.gov.uk/UKThreatLevel/UKThreatLevel.xml

In the data selection, terrorist incidents and natural hazard events have been selected as demonstration data sources for open source physical security data. These sources are used as examples from the larger world of all possible physical security data, which can also include additional data sources (e.g. information feeds concerning product recalls of non-cyber goods).

4.2.3 Development and Demonstration Data

For development and demonstration purposes, it has been desirable to have an easily controllable development and demonstration environment available. Two such development and demonstration activities are relevant in Task 4.2.

First, it has been useful to work from a small-scale world model in order to gradually expand testing until the system crosses over to fully specified and modelled demonstration and real environments. To this end, “Lummerland Rail” has been created as a toy example railway for data modelling and technical testing purposes

[30] All web resources accessed 27.03.2022.

in T4.2. While the name Lummerland Rail references a children's book series[31], the toy example railway used in T4.2 uses modern technology and devices. A summary of these data is presented in Annex II of this deliverable document.

Second, demonstration data from the end user partners involved in the project was integrated for validation and demonstration purposes. These data do not necessarily represent devices and system deployed by operators in the real world but are representative and sufficiently similar to real-world deployments to be suitable for realistic validation and demonstration activities. These data are documented in the relevant WP8 deliverables in order to ensure that they are reviewed and redacted as appropriate in order to safeguard real or realistic end user partner system configurations.

[31] See https://en.wikipedia.org/wiki/Jim_Button_and_Luke_the_Engine_Driver (accessed 27.03.2022) for further information on the source for the naming of the railway toy example.

5. Conclusion

This section presents a brief overall summary of the content of this deliverable and reproduces the summary capabilities matrix that indicates which of the functionalities identified and described throughout this deliverable address which of the required functionalities that have been outlined in Section 2.

5.1 Summary

This deliverable has provided an overview of the approach to cyber-physical threat detection in terms of functionalities and components that have been implemented in order to integrate the envisioned OSINT approach to data acquisition to identify cyber-physical threats in the railway domain. The general system architecture for the identified components as well as the implementations created were described. The approach to data representation, custom MISP data model implementations and main issues concerning data sources and data acquisition were discussed.

The information provided in this document represents the final status of development in Task 4.2 of SAFETY4RAILS, completed in the first half of the second year of development in the project. The results of this work will be used in other tasks of the project for validation, integration and will be used beyond the lifetime of the project as presented in the exploitation plan for the project. Of particular interest for the latter is the modular construction of the OSINT system, which allows the participating partners to deploy their contributions independently of those of other partners in the task if they so desire.

5.2 Capability Matrix

In order to review in which way the components proposed for development contribute to the identified relevant system functionalities that the OSINT system in SAFETY4RAILS should provide, the table below relates the required functionalities to components in a capability matrix.

In the table functionalities and components respectively are identified by their unique IDs and briefly characterised by (shortened) summary texts intended to help with recalling the respective functionalities and component characterisations. In each component row, an “x” signifies that a component contributes towards enabling a functionality either partially or completely. As outlined in Section 2, the developed components have addressed all requirements except for DM-MSP-01, which was found to no longer be required for use in the project during development.

TABLE 11: CAPABILITIES MATRIX RELATING SUGGESTED FUNCTIONALTIIES AND SUGGESTED COMPONENTS

	FUNC-DA-01 Retrieval from cyber security data sources	FUNC-DA-02 Identification of vulnerable & compromised assets	FUNC-DA-03 Retrieval from physical security data sources	FUNC-PP-01 Parsing of standard format data feeds	FUNC-PP-02 Analysis of unstructured data sources	FUNC-SR-01 Operation of a suitable database system	FUNC-DS-01 Analysis of newly added data for threat intelligence	FUNC-DS-02 Generation of statistics to identify trends and ranks	FUNC-DN-01 Exposing data access functionalities	FUNC-DM-02 Messaging data updates to recipient components
DA-TIS-01 Internet exposed asset crawlers	X									
DA-TIS-02 Vulnerability and exploit scanning crawlers		X								
DA-TIS-03 Malware repository crawlers with Yara rules	X	X		X						
DA-TIS-04 Malware social media crawlers with Yara rules	X				X					
DA-TIS-05 Phishing campaign monitoring crawlers		X								
DA-TIS-06 Threat intel feed selection and configuration	X			X		X				
DA-MSP-01 Threat intel feed selection and configuration for cyber threats	X			X		X				
DA-MSP-02 Structured data source integration for physical threat data			X	X						
DA-MSP-03 Social media data source integration for physical threat data			X							
PA-MSP-01 Threat intel feed automated relevance filtering				X	X					
PA-MSP-02 Structured physical threat automated data feed filtering				X	X					

PA-MSP-03 Semantic social media data parsing for physical threat detection					X					
SR-TIS-01 Data repository operation						X				
SR-TIS-02 Data model customisation						X				
SR-MSP-01 Data repository operation						X				
SR-MSP-02 Data model customisation						X				
DS-MSP-01 Rule evaluation over database							X			
DS-MSP-02 Generation of summary statistics								X		
DM-TIS-01 Synchronisation with SAFETY4RAILS MISP instance						X				
DM-MSP-01 Custom API endpoint development[32]									X	
DM-MSP-02 Message broker integration										X

[32] No custom API endpoint has been developed as no external system required custom API endpoint functionalities and preferred to access data via the SAFETY4RAILS messaging system.

REFERENCES

1. **Steele, R.D.** Open Source Intelligence. [book auth.] L.K. Johnson. *Handbook of Intelligence Studies.* New York : Routledge, 2007, pp. 129-147.

2. **Wikimedia Foundation.** Open-Source Intelligence. *Wikipedia.* [Online] [Cited: 19 04 2021.] https://en.wikipedia.org/wiki/Open-source_intelligence.

3. **Lowenthal, M.M.** Open-Source Intelligence: New Myths, New Realities. [book auth.] R.Z. George and R.D. Kline. *Intelligence and the National Security Strategist: Enduring Issues and Challenges.* Lanham : Rowman & Littlefield, 2004, pp. 273-278.

4. **Free Software Foundation.** GNU Affero General Public License v3.0 or later. [Online] [Cited: 19 04 2021.] https://spdx.org/licenses/AGPL-3.0-or-later.html.

5. **Gartner.** Security Threat Intelligence Products and Services Reviews and Ratings. *Gartner.* [Online] [Cited: 19 04 2021.] https://www.gartner.com/reviews/market/security-threat-intelligence-services.

6. **Wagner, C., et al.** MISP: The Design and Implementation of a Collaborative Threat Intelligence Sharing Platform. *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security.* Vienna : ACM, 2016, pp. 49-56.

7. **Kellet, M. and Bernier, M.** *Cyber Threat Data Model. High-Level Model and Use Cases.* s.l. : Defense Research and Development Canada, 2016. DRDC-RDDC-2016-D080.

8. **Magar, A.** *State-of-the-Art in Cyber Threat Models and Methodologies.* 2016. DRDC-RDDC-2016-C132.

9. *Distributed Security Framework for Reliable Threat Intelligence Sharing.* **Preuveneers, D., et al.** 2020, Security and Communication Networks, pp. 1-15.

10. **MISP Project.** Best Practices in Threat Intelligence. *MISP Project.* [Online] [Cited: 19 04 2021.] https://www.misp-project.org/best-practices-in-threat-intelligence.html.

ANNEXES

The annexes to this deliverable contain auxiliary information. Annex I. “Glossary and Acronyms” contains an overview of abbreviations and acronyms used in this deliverable.

ANNEX I. GLOSSARY AND ACRONYMS

The following table lists and defines or describes the abbreviations and acronyms used in this deliverable.

TABLE 12: GLOSSARY AND ACRONYMS

Term	*Definition/Description*
AWS	Amazon Web Services
DoA	Description of Action
IDS	Intrusion Detection System
IoC	Indicator of Compromise
IT	Information Technology
JSON	JavaScript Object Notation
MISP	Malware Information Sharing Platform
OSINT	Open Source Intelligence
OT	Operational Technology
RSS	Rich Site Summary, a data feed format
SCADA	Supervisory Control and Data Acquisition
SOC	Security Operations Centre
TI	Threat Intelligence
TTP	Tactics, Techniques and Procedures
VPC	Virtual Private Cloud

ANNEX II. OVERVIEW OF THE TESTING AND DEMONSTRATION DATA TOY EXAMPLE

This annex describes the toy example that was used for initially testing and then transitioning the data used for development of the OSINT subsystem to the final demonstration and evaluation scenarios within SAFETY4RAILS Task 4.2.

Toy Example Overview

While fundamentally the toy example is an abstract listing of components, we frame the toy example setup in order to make it easier for developers to discuss it and for data contributors to provide data that can be resolved to locations and environments instead of abstract inventories.

The example that we employ is based on the Morrowland ("Lummerland") railway written by German children's author Michael Ende (best known for being the author of the literary source for the movie "The Never Ending Story"). The following images depict the railway setup from a puppet TV show adaptation and as provided by a railway hobbyist:

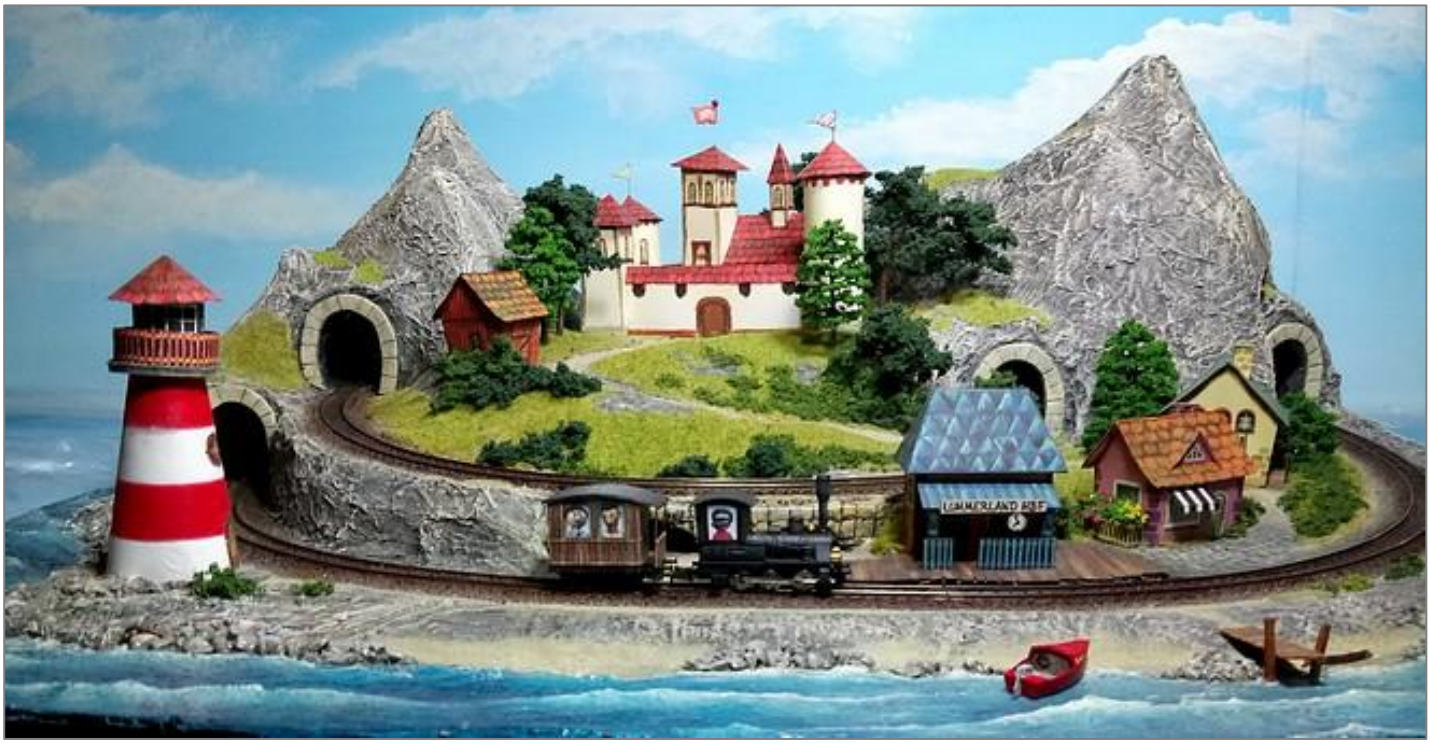


FIGURE 30: RAILWAY AS DEPICTED IN A TELEVISION PROGRAMME

For the purpose of the project, the components selected for our version of this railway system are selected from present-day equipment. The railway is fully electrified and includes the following structures and properties that can be used for adding suitable components:

- A single main train station (lower right centre of the map)
- A single additional train station (lower right of the map)
- A train depot (centre of the map)
- Four single-track tunnels (left and right borders of the map)
- A largely single-track railway environment that requires numerous signals
- A town centre area
- A level crossing (upper right area of the map)

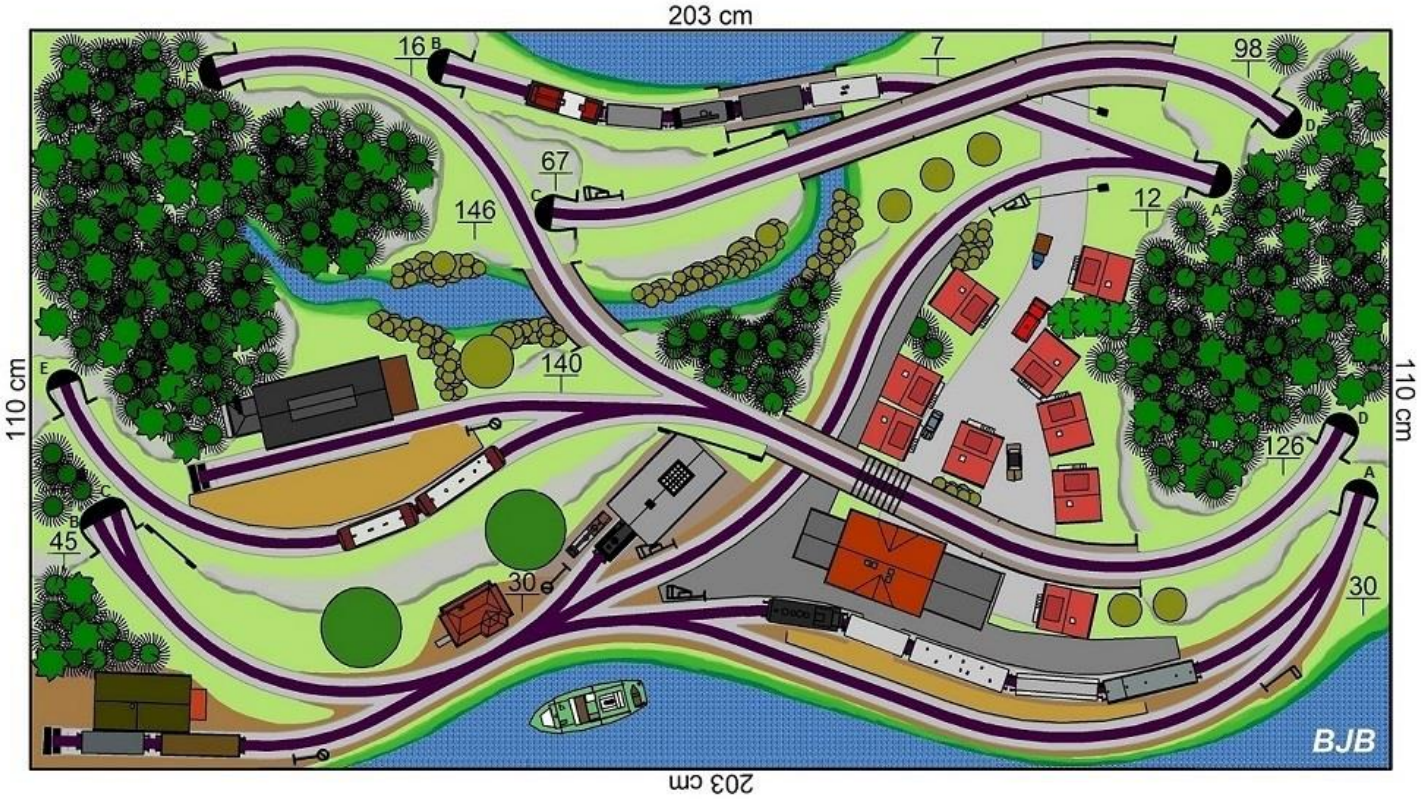


FIGURE 31: RAILWAY SYSTEM OVERVIEW USED FOR THE TOY EXAMPLE

Component Table

This subsection presents a table listing the components deployed as part of the toy example railroad. We use separate tables for specific types of locations such as the main train station of the island.

TABLE 13: TRAIN STATION EQUIPMENT

Item	Location(s)	Purpose(s)	Notes
6 Dell Optiplex office PCs, Windows 10 Professional operating system, Microsoft Office	Train station offices, station operator room	General office use	Office PCs are connected to the train station wifi network
20 Linksys LAPAC1300CE wireless access points	Train station real estate	General office station networking	Centrally cloud managed by railway operator
4 Linksys LGS352C-EU port managed ethernet switches	Technical backroom	General network connectivity for all station requirements	Centrally cloud managed by railway operator
Linksys Cloud Manager software	Online	Management of train station network infrastructure	Directly manages the networking infrastructure on location

Novisign Cloud Digital Signage system	Online	Management of in-station displays	Directly manages the displays on location
10 PPDS Philips D-Line 34BDL4031D/00 displays	Indoor and trackside at train station	Information display	Managed via Novisign Cloud Digital Signage system
Linksys Network video recorder SKU LNR-0208C-AP	Technical backroom	CCTV data recording, storage and live display	Connects to station operator room for live monitoring
6 Linksys LCAD03FLN indoor dome camera	Train station indoor public areas	CCTV data acquisition	
12 Linksys LCAD03-VLNOD-AP	Train station outdoor public areas	CCTV data acquisition	
Siemens SITRAS SCS-AC station control system	Station power control room		
Siemens Sitras RSC remote control system	Cloud	Electrification station control system	Monitored centrally and in the station control, station operator room
2 Siemens high-voltage transformer panels	Station trackside	Electrification system	
4 Siemens contact line feeder panels	Station trackside	Electrification system	
4 Siemens incoming panels	Station trackside	Electrification system	
2 Siemens Sitras SCS-FFP frame fault protection units	Station trackside	Electrification system	
Siemens Sitras SCS-TTU transfer trip unit	Station trackside	Electrification system	
1 Siemens Trainguard LZG 700 M	Station operator room	Automatic train control system	
2 Siemens Trainguard IMU 100	Inductive message transmission system for train data transmission	Message transmission through rail electric transmission	

TABLE 14: TRACKSIDE EQUIPMENT

Item	Location(s)	Purpose(s)	Notes
Siemens Trainguard MT Zub trackside control system	Trackside coupled with signalling components	Train control trackside unit	

TABLE 15: ONBOARD EQUIPMENT

Item	Location(s)	Purpose(s)	Notes
Siemens Trainguard MT Zub on-board control system	Technical operating cabinet	Train control on-board unit	

SAFETY4RAILS

Partners:

Fraunhofer EMI
UIC
Metro de Madrid
EGO Genel Müdürlüğü
RFI Rete Ferroviaria Italiana – Gruppo Ferrovie dello Stato Italiane
Leonardo
ceis avisa partners
STAM Mastering Excellence
IC information company
RMIT University
University of Reading
etra I+D
Demokritos National Centre for Scientific Research
RINA
WINGS ICT Solutions
Newcastle University
EOS European Organisation for Security
Innova Integra
LAUREA Ammattikorkeakoulu University of Applied Sciences
CybeServices Made in Europe
erGE
MTRS
INTRACOM TELECOM
Universitas Miguel Hernández
TREE Technology
Elbit Systems C4I and Cyber
ProRail
Comune di Milano

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 883532.