Co-funded by the European Union

ECCC

Ref. Ares(2024)6852430 - 27/09/2024

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

CYDERCO - CYber DEtection, Response and COllaboration

D2.3 CYDERCO Platform Design - Initial Version

Deliverable date: 2024-09-26

Status: Final

Version: 1.0

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

List of changes

Version	Date	Description	Author(s)
0.1	10.05.2024	First Version	Cristian Radu, Ioana-Andreea Craciun, Mircea Avram
0.2	26.09.2024	Final Version	Cristian Radu, Ioana-Andreea Craciun, Mircea Avram
1.0	26.09.2024	Quality Assurance Review	Liana Predut, Ioana-Andreea Craciun

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Contributors

Role	Contributor's Name	Entity Name - Beneficiary
Deliverable Lead	Cristian Radu	Eviden Technologies
Contributor	Ioana-Andreea Craciun	Eviden Technologies
Contributor	Mircea Avram	Eviden Technologies
Contributor	Gabriel Petre	Eviden Technologies
Contributor	Andrei Chipaila	Eviden Technologies
Contributor	Alexandru Rusandu	Eviden Technologies
Contributor	Alexandru Velcea	Eviden Technologies
Contributor	Dan Brasov	Eviden Technologies
Contributor	Stefan Iordache	Eviden Technologies
Contributor	Felician Moldovan	Eviden Technologies
Contributor	Elena Rusu	Eviden Technologies
Contributor	Elena Stanciu	Eviden Technologies
Contributor	Eva Maia	ISEP
Contributor	João Vitorino	ISEP
Contributor	Isabel Praça	ISEP
Contributor	Rodrigo Diaz	ATOS
Contributor	Miguel Martin Perez	ATOS
Contributor	Mihai Nena	DNSC
Contributor	Stefan Tanase	DNSC
Contributor	Gabriel Ene	DNSC

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Approvers

Entity Name - Beneficiary	Project Manager	Signature
Eviden Technologies SRL	Ovidiu Calancea	X Ovidiu Calancea Digitally signed by Ovidiu Calancea Date: 2024.09.27 15:12:49 +03'00'
Instituto Superior De Engenharia Do Porto	Isabel Praça	X Assinado por: Isabel Cecília Correia da Silva Praça Gomes Pereira Identificação: BI09480460 Data: 2024-09-27 às 12:14:09
Directoratul National De Securitate Cibernetica	Christine Demeter	X Semnatar: Demeter Christine Data si ora semnarii: 2024.09.27 09:55:08 +0300
Atos Spain SA	Rodrigo Diaz Rodriguez	Rodrigo Diaz Rodriguez X Digitally signed by Rodrigo Diaz Rodriguez Date: 2024.09.26 17:51:07 +02'00'

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Contents

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

1. Glossary: Acronyms, Terms and Abbreviations

1.1 Acronyms

EU	European Union
DA	Data Analytics
TIP	Threat Intelligence Platform
CTI	Cyber Threat Intelligence
DR Hub	Detection and Response Hub
HIDS	Host Intrusion Detection System
NTA	Network Traffic Analysis
ETL	Extract Transform Load
FIM	File Integrity Monitoring

1.2 Abbreviations

Req.	Requirements
SIG	Signatures
N-DATA	Network data
E-DATA	Endpoint data
ID	Identifier
Int	Interface
Tech	Technical
Data FWD	Data Forwarder
Misc	Miscellaneous
Transf	Transformation

2. Introduction

This High-Level Design document outlines the architectural specifications of CYDERCO and serves as a reference for design information and project components.

Establishing a high-level design is paramount in developing new systems. The solution's architecture is the backbone of the entire CYDERCO infrastructure, homogeneously integrating various layers, views, technical specifications, and associated operational objectives in a comprehensive view.

There are two versions of this deliverable:

- "D2.3 - CYDERCO Platform design – Initial version" linked to task T2.3 in Work Package 2
- "D2.4 - CYDERCO Platform design – Final version" also linked to task T2.3 in Work Package 2

The task focuses on defining the platform's high-level architecture and identifying key configuration items such as software and hardware equipment. It builds upon previous tasks in the same work package that addressed functional requirements and stakeholder analysis elicitation. Manual operations will be subsequently identified based on these elements.

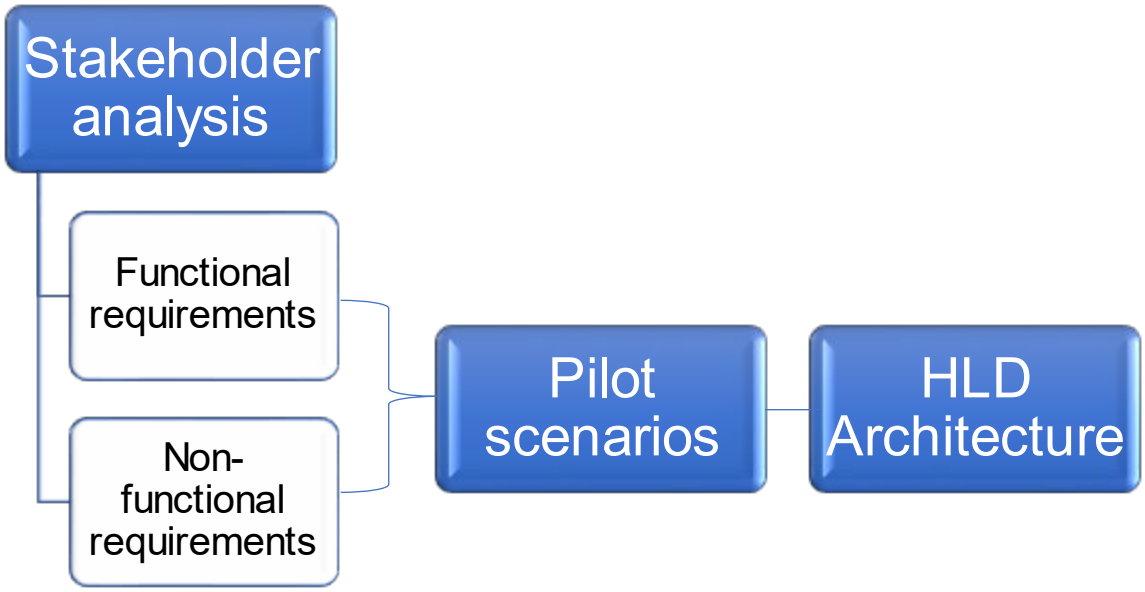

Figure 1 - WP2 tasks supporting CYDERCO HLD Architecture

This deliverable aims to provide a holistic view of the platform's components, interfaces, stakeholders, and associated operations, using graphical representations and detailed textual input.

2.1 General Information

Despite incident response mechanisms being constantly updated, infrastructure systems struggle to keep up with the rampant proliferation and sophistication of cyberattacks.

SOC ecosystems must embrace disruptive technologies and cross-border collaboration to pick up the pace and ultimately stay ahead. This is precisely what the call DIGITAL-ECCC-2022-CYBER-03-SOC aims to tackle.

The deliverable D2.3 is part of a wider project called CYDERCO, whose purpose is established early on. CYDERCO aims to strengthen cybersecurity resilience by enhancing the detection and response capabilities of SOCs and other relevant entities involved in the cybersecurity ecosystem with disruptive technologies, streamlined threat intelligence sharing, and fine-tuned collaboration mechanisms.

To this effect, CYDERCO will deliver a comprehensive platform serving the two underlying objectives. First, we will develop a platform (from this point referred to as "Detection and Response Hub") to enable advanced detection of malicious activities.

The second objective coincides with the emergence of a different component (from this point referred to as "Threat Intelligence Platform"). Each covers different operational needs and is designed to help mitigate different challenges.

The Detection and Response Hub will enable interested entities to perform advanced cyber threat detection and analysis with real-time monitoring. It consists of 4 modules, each developed under different tasks and work packages throughout the project:

- Data Analytics: Work Package 3, task T3.1
- Network Traffic Analysis (NTA): Work Package 3, task T3.2
- Host Intrusion Detection System (HIDS): Work Package 3, task T3.3
- AI-driven Analytics: Work Package 3, task T3.4

On the other hand, the Threat Intelligence Platform will streamline threat intelligence sharing among entities to increase trust and improve reaction time when dealing with cyber incidents through actionable, relevant, and high-confidence CTI.

It consists of 4 modules, each developed under different tasks and work packages throughout the project:

- TI Collection and Storing: Work Package 4, task T4.1
- TI Sharing: Work Package 4, task T4.2
- TI Enrichment: Work Package 4, task T4.3
- Sighting support: Work Package 4, task T4.4

3. Methodology

Faithfully embodying our holistic approach, deliverable no D2.3 developed under the current T2.3 tasks builds upon the results of previous work and serves as a foundation for future tasks. Thus, the architecture and technical design of the platform developed in this deliverable will guide the iterative development process to be carried out in WP3, WP4, and WP5.

An integral part of Work Package 2, the platform design leverages the outcomes of previous tasks and deliverables developed under the same work package, as follows:

- Stakeholders' analysis - task no T2.1, deliverable no D2.1
- Requirements Elicitation - task no T2.1, deliverable no D2.1
- Pilot definition for platform validation - task no T2.2, deliverable no D2.2

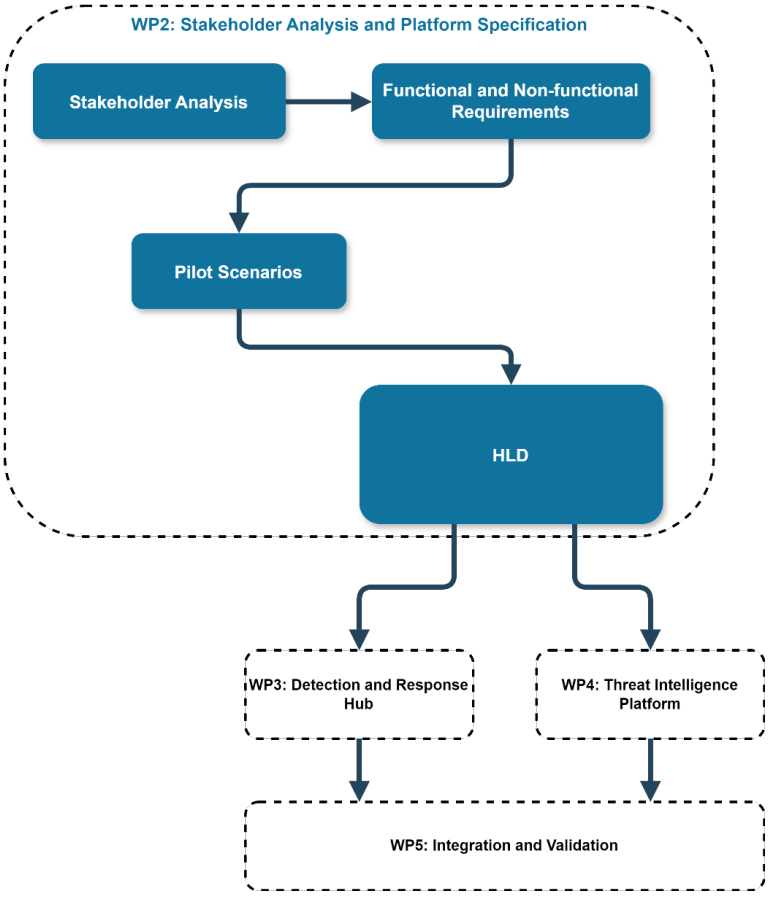

Figure 2 - Flow view of WP2 tasks and WP3, WP4, and WP5

This chapter provides insight into the development of the platform design and the reasoning behind the architectural decisions. We outline the approach and procedures used in this documentation, including the structure, components, interfaces, building blocks, key features, and requirements.

We also discuss the applicability of operations at different levels and the integration of application, data, and technology layers to give a thorough overview of the methodology for reproducibility and reliability. Our working methodology is centered on key concepts that guide our process, as follows:

Architectural Requirements and Constraints

In this section, we describe the architectural principles that guide the design and future implementation of the system, including the design patterns, standards, and best practices along with the rationale behind certain design decisions and available alternatives (where applicable).

As part of this step, we approached the architecture's compliance with industry policies, standards, frameworks, and guidelines (such as Agile, and DevOps).

We honed the work developed during previous tasks such as use cases, scenarios, key features, and requirements, which allowed us to identify the dependencies for the design activity.

Supported Policies, Processes, Guidelines, and Standards

At CYDERCO, we understand the critical importance of adhering to established frameworks in our platform design. Thus, the design is focused on compliance and alignment with a comprehensive array of European Union (EU) cybersecurity policies, directives, and standards.

We integrate foundational directives such as NIS 2 and DORA, critical infrastructure protection frameworks like CER, and key guidelines from ENISA.

By incorporating best practices for threat intelligence and incident response, CYDERCO aims to meet regulatory requirements and enhance organizations' overall security posture across Europe.

Our commitment contributes to improved cybersecurity resilience, operational continuity, and cross-border collaboration.

High-Level Architecture

The overview includes a general presentation of the platform's structure and underlying components. This is supported by a visual representation (*Figure 5 - High-level view on CYDERCO solution components*).

This methodology phase builds upon previous design choices and refines them considering any prerequisites necessary for the architecture, such as infrastructure, technologies, and corresponding constraints.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Information System and Technology Architecture

Finally, the system's architectural design provides a high-level overview of the proposed solution architecture, including the system components, interfaces, and interactions.

We present different views (conceptual, logical, physical) of the system's design supported by various diagrams as visual facilitators.

The architectural representation spreads across four layers:

- High-level architecture (described in chapter 5)
- Application (described in chapter 6.1)
- Data (described in chapter 6.2)
- Technology (described in chapter 6.3)

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

4. Requirements and Constraints

4.1 Architecture Principles

This chapter will outline the guiding principles behind the design and implementation of the CYDERCO Platform. It will cover the system design, purpose, decisions made during the planning phase and their consequences (benefits, respectively drawbacks of the chosen architecture if any), alternative options considered (if any), and the logic behind the architecture choices.

The below table illustrates the underlying format of each principle stated in this chapter.

Table 1 - Format adopted to showcase the architecture principles

Name	The title of the principle is summarized in a brief and memorable phrase or word.
Purpose	Offers a brief overview of the main goals and practical uses of the principle. It should be detailed enough to convey the core concept of the principle while remaining concise.
Logic	Will highlight the benefits (and drawbacks, if any) of specific architectural choices and detail their primary uses within the operational realm. It will present a structured rationale for specific design and architecture decisions, considering functionalities, technologies, and other relevant factors.
Consequences	Will highlight the requirements incorporated in the architectural view and the implications associated with the adoption of the principle in the development of the solution.

To effectively govern the architecture development process, our principles reflect a consensus across consortium members, relying on the following pillars:

- **Consistent** - the principle should take precedence in supporting the resolution of contradictory or controversial situations (for instance, clashing features such as security versus usability or accessibility).
- **Comprehensive** - the principles should be easy to understand, unambiguous, and precise.
- **Stable yet flexible** - while principles are meant to be enduring for consistency purposes, evolutionary needs must be accommodated based on a traceable process to add/remove/modify principles and keep them relevant, updated, and reliable.

To implement different system features, the design decisions are carefully made based on the analysis of Make, Buy, or Re-use. While a dedicated budget is allocated for purchasing hardware

equipment, software components, and their associated interfacing allow partners to showcase their specific contributions and demonstrate the Make and Re-use components.

The design alternatives consider various aspects related to the platform components, including native capabilities such as SIEM and SOAR, open-source options for customization and preferential development, seamless integration interfaces, and more.

Considering all these factors, the final technology options we have chosen are as follows:

- Software components: External SaaS (GitLab), Hypervisor Proxmox VE, Microservices
- Hardware components: Compute, Storage, Network

The development of this high-level design documentation is guided by design principles that encompass key concepts such as *usability, security, reliability, availability, performance, scalability, portability,* and *reusability*.

Below are our primary architecture principles:

1. Design and develop with functional perspective
2. Interoperability
3. Improved accessibility and usability
4. Continuous innovation

4.1.1 Principle 1

- **Name: Design and develop with functional perspective**
- **Purpose:**

Our first principle is tightly connected to the third one, highlighting usability and accessibility.

We prioritize user-centered design, recognizing that functionality is crucial in developing a solution that provides finely tuned and tailored mechanisms for its users. In this case, we focus on detection, analysis, incident response, and threat intelligence capabilities.

- **Logic:**

The reasoning behind this architectural decision implies that the CYDERCO platform design fulfills its intended purpose. It delivers a solution that performs the intended functions in a consistent, reliable, easy, and safe-to-use manner.

Another argument favoring this principle, in addition to **practicality**, is effectiveness. Thus, we aim to develop a platform that meets user needs and precisely addresses their functional expectations, without generating unnecessary clutter or excessively intricated modules that don't contribute to the solution's core functionality.

- **Consequences:**

Functionality can be boiled down to **effectiveness**, which implies alignment with users' or stakeholders' expectations, needs, and requirements.

This entails considering the users' needs and the specific context in which the platform will be used.

To this effect, we will develop a **straightforward and intuitive** graphical user interface, offering centralized management and extended visibility into operational data that security personnel such as SOCs and CSIRTS can incorporate into their activities.

The feedback gathered from deliverable D2.1 illustrating the **end users' perspective** allowed us to tailor the solution based on their specific requirements related to *accessibility, understandable data access, tracking change in data sources, query languages, lookup times*, and more.

4.1.2 Principle 2

- **Name: Interoperability**
- **Purpose:**

Interoperability is a core principle guiding our platform design. By adhering to standards, we ensure consistency, improve system management, enhance user satisfaction, and reduce costs. Interoperability standards also enable support from multiple vendors and facilitate integration.

- **Logic:**

Our platform's collaborative nature requires a modular design for **easy integration with other systems and future extensions**. Alignment with applicable standards is a mandatory design criterion to ensure interoperability with existing and future components.

CYDERCO's platform securely shares and analyzes large datasets, enhancing TI availability, usability, and interoperability among entities. Our solution supports the rapid deployment of signatures and countermeasures across various entities, leveraging well-known TI standards for high interoperability with security controls.

- **Consequences:**

Utilization of standard data formats in cybersecurity is paramount. For instance, threat intelligence shall be exported in widely supported industry formats such as STIX, JSON, MISP, and CSV.

To facilitate the ingestion of information by other security controls for detection or blocking purposes, the Threat Intelligence Sharing component will support the exchange of information regarding the Course of Actions in standardized MISP and STIX formats.

Extended OS support is also considered. For instance, the HIDS module will provide the core functionality for Windows and enable support in other OSs such as Linux or Mac using cross-platform technologies and programming languages.

4.1.3 Principle 3

- **Name: Improved accessibility and usability**
- **Purpose:**

In addition to standardization and cross-platform compatibility, we have focused on improving the usability and accessibility of our applications. The CYDERCO platform will feature an intuitive, **user-friendly interface** that allows for **easy navigation** and **centralized visibility** across different entities.

- **Logic:**

The rationale behind this architectural choice is that ease of use is a positive incentive encouraging users to work within the integrated information environment instead of relying on isolated systems for different tasks. This approach reduces the need for extensive training and technical skills while improving decision-making and response times for critical incidents.

- **Consequences:**

CYDERCO will provide a unified dashboard that can generate **various technical and non-technical reports**, graphs, and other **visual documentation**. This ensures that the displayed information is sufficiently versatile to meet the diverse functional needs of users.

4.1.4 Principle 4

- **Name: Continuous Innovation**
- **Purpose**:

Our design decisions are based on the options of Make, Buy, or Re-use and building upon existing solutions to deliver functional and mature products with minimal risk. However, we also recognize the importance of generating unique differentiators and crafting intellectual property that drives beneficial innovation.

In addition to reusing and replicating solutions, we are committed to continuous innovation and improving the efficiency of cybersecurity operations. We achieve this by delivering value through disruptive technologies, state-of-the-art developments, and streamlined collaboration mechanisms.

- **Logic:**

Threat actors are constantly pushing boundaries and trying to expand the attack surface with new techniques. To address these challenges, we aim to develop and adopt disruptive technologies that fortify cyber resilience and increase situational awareness.

- **Consequences:**

By embracing the principle of continuous innovation, CYDERCO fosters an environment of perpetual evolution that scales with operational needs.

We aim to strengthen detection, analysis, incident response, and threat intelligence sharing capabilities with state-of-the-art technology in areas such as AI. Consequently, CYDERCO will incorporate AI-based techniques throughout the detection and response phases and in network security to proactively identify and prevent cyber threats. We are committed to contributing to the innovation of AI in the cybersecurity field.

4.2 Supported policies, processes, guidelines, and standards

The CYDERCO platform design is aligned with a wide range of European Union (EU) cybersecurity policies, directives, and standards. Recognizing the importance of adhering to established frameworks, CYDERCO aims to meet regulatory requirements and support broader

EU objectives for enhancing cybersecurity resilience, operational continuity, and cross-border collaboration.

This section outlines the key policies, processes, guidelines, and standards that CYDERCO supports, ensuring that the platform contributes effectively to the overall cybersecurity landscape in the EU. These include foundational cybersecurity directives, such as NIS 2 and DORA, critical infrastructure protection frameworks like CER, as well as important guidelines issued by ENISA and standards from ISO.

Furthermore, CYDERCO integrates best practices for threat intelligence and incident response lifecycles, ensuring that it operates within the best possible frameworks to support both national and EU-wide cybersecurity initiatives.

CYDERCO can provide cybersecurity solutions that are aligned with both current and emerging regulatory demands.

4.2.1 Alignment with Key EU Policies and Directives

CYDERCO has the capabilities to align with the cybersecurity directives and regulations in the European Union, ensuring that it not only meets regulatory requirements but also strengthens the security posture of organizations across Europe.

One of the foundational frameworks that CYDERCO supports is the **NIS 2 Directive**. This directive is important for enhancing the security of network and information systems across the EU, particularly within sectors that are essential to societal functioning.

By integrating advanced threat detection and response capabilities, CYDERCO ensures that organizations can swiftly identify and react to cyber threats. Furthermore, the platform's focus on information sharing aligns seamlessly with NIS 2's mandate for enhanced collaboration between EU member states, helping to create a more secure and resilient digital environment.

In addition to NIS 2, the **CER Directive** plays a pivotal role in protecting critical infrastructure across Europe. CYDERCO can meet the needs of IT enclaves operating within critical sectors such as energy, transport, and healthcare.

The platform's ability to provide continuous monitoring and proactive threat mitigation ensures that critical entities can maintain operational continuity even in the face of sophisticated cyberattacks, thus supporting the resilience objectives of the CER Directive.

Financial institutions, which are increasingly targeted by cyber threats, benefit significantly from CYDERCO's alignment with the **Digital Operational Resilience Act (DORA)** (European

Insurance and Occupational Pensions Agency, DORA, 2024). This act requires financial entities to ensure operational resilience, and CYDERCO supports this by offering tools that enable rapid detection and effective response to incidents. By doing so, CYDERCO helps financial institutions not only comply with DORA but also safeguard the integrity and availability of their services.

4.2.2 Supporting EU Guidelines and Best Practices

Beyond compliance with directives, CYDERCO also incorporates best practices and guidelines issued by the **European Union Agency for Cybersecurity (ENISA)** (ENISA, 2024). ENISA's recommendations on incident reporting and Security Operations Centre (SOC) best practices are embedded within CYDERCO's design.

By following these guidelines, CYDERCO ensures that users have access to streamlined processes for detecting and reporting incidents, while upholding the highest standards of cybersecurity management across the EU.

4.2.3 Integration with Cybersecurity Standards and Frameworks

CYDERCO's commitment to security is further demonstrated by its alignment with internationally recognized standards, such as **ISO/IEC 27001** and **ISO/IEC 27005**. These standards form the foundation of effective information security management and risk assessment practices.

By aligning with these frameworks, CYDERCO ensures that organizations can manage their security risks efficiently while maintaining the highest levels of data protection. This not only enhances trust in the platform but also ensures that it can be integrated into the existing security infrastructures of organizations across Europe.

4.3 Key features and requirements

The current high-level design leverages work performed during previous tasks to deliver a cohesive and structured overview of the platform's components, layout, and supporting infrastructure.

The requirements - developed to reflect the actual needs and expectations of potential end-users - are an output of task T2.1 "Stakeholders' analysis & requirements elicitation".

The following tables present the functional requirements gathered from the stakeholders' perspective and correlated with the work package, task number, and component.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Table 2 – Functional requirements mapped to WP, Task, Component in the Detection and Response Hub

Detection and Response Hub (Work Package 3)							
Task T3.1		**Task T3.2**		**Task T3.3**		**Task T3.4**	
Data Analytics	Functional Req.	AI Analytics	Functional Req.	NTA	Functional Req.	HIDS	Functional Req.
	Data correlation		Data preprocessing		Traffic monitoring		Anomaly-based detection from host behavioral patterns
	Data visualization		Smart correlation		Anomaly detection		Alert forwarding
	Intuitive GUI				Alert generation		AI-based anomaly detection
	Alert forwarding				Alert formats		Outgoing network activity modelling
	Role-based access control				Generate signature-based alerts (from IoCs)		Incoming network activity modelling
	Technology support						Automated configuration
	Sighting support						Resource consumption anomaly detection
	Ti enrichment						Pattern correlation
Data Forwarder	Data ingestion						Extended pattern correlation
Data Forwarder	Normalization						Log analysis
Data Forwarder	Data parsing						Detection capabilities - Categories
Data Forwarder	Data enrichment						

The component field value from the above table represents the technical component as it will be described in the Application, Data, and Technology Architectures (Chapter 6 in the document herein).

Table 3 - *Functional requirements mapped to WP, Task, Component in the Threat Intelligence Platform*

Threat Intelligence Platform (Work Package 4)							
Task T4.1		**Task T4.2**		**Task T4.3**		**Task T4.4**	
TI Collection & Storing	Functional Req.	TI Sharing	Functional Req.	TI Enrichment & Contextualization	Functional Req.	Sighting	Functional Req.
	Ingestion support for 3rd party TI feeds		Field encryption		Dashboard enrichment and visualizations		Threat intelligence support
	Add/Remove/ Edit events		Field anonymization		Asset information retrieval		
	TI ingestion from DR Hub		Course of actions – TI sharing		Confidence level		
	IoC datasets		Manual export format		MITRE support		
	TI sources		Information synchronization		Automated enrichment		

*Note 1: The above tables showcase only the functional requirements without which the project could not be implemented. Non-functional features describing performance, scalability, etc. are not included in this presentation.

*Note 2: A more comprehensive, but not exhaustive, list of requirements including functional, non-functional, mandatory, as well as medium, or low-priority requirements is provided in deliverable "*D2.1. Stakeholders' analysis & requirements elicitation"* for further reference.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

4.4 Services created

As mentioned earlier in the Methodology, the current high-level design builds on previous tasks. The scenarios developed to validate the solution and its capabilities are an output of task T2.2 "Pilot Definition".

Below are the two scenarios we created to demonstrate CYDERCO's effectiveness in real-world cybersecurity situations:

- **Scenario 1: Data Exfiltration on National CERT**

This scenario simulates a National CERT environment with a virtualized IT&C network and procedural frameworks.

- **Scenario 2: Ransomware Attack on a Public Hospital**

This scenario replicates a virtual hospital IT environment with real-world constraints and challenges.

Both scenarios follow industry standards and best practices such as NIST 800-61 and SANS Incident Response lifecycle phases.

In Figure 3 below, we highlighted the sequence of events corresponding to each incident response phase in the first scenario (Preparation, Detection, Containment, Eradication, Recovery, and Post-Incident Activities).

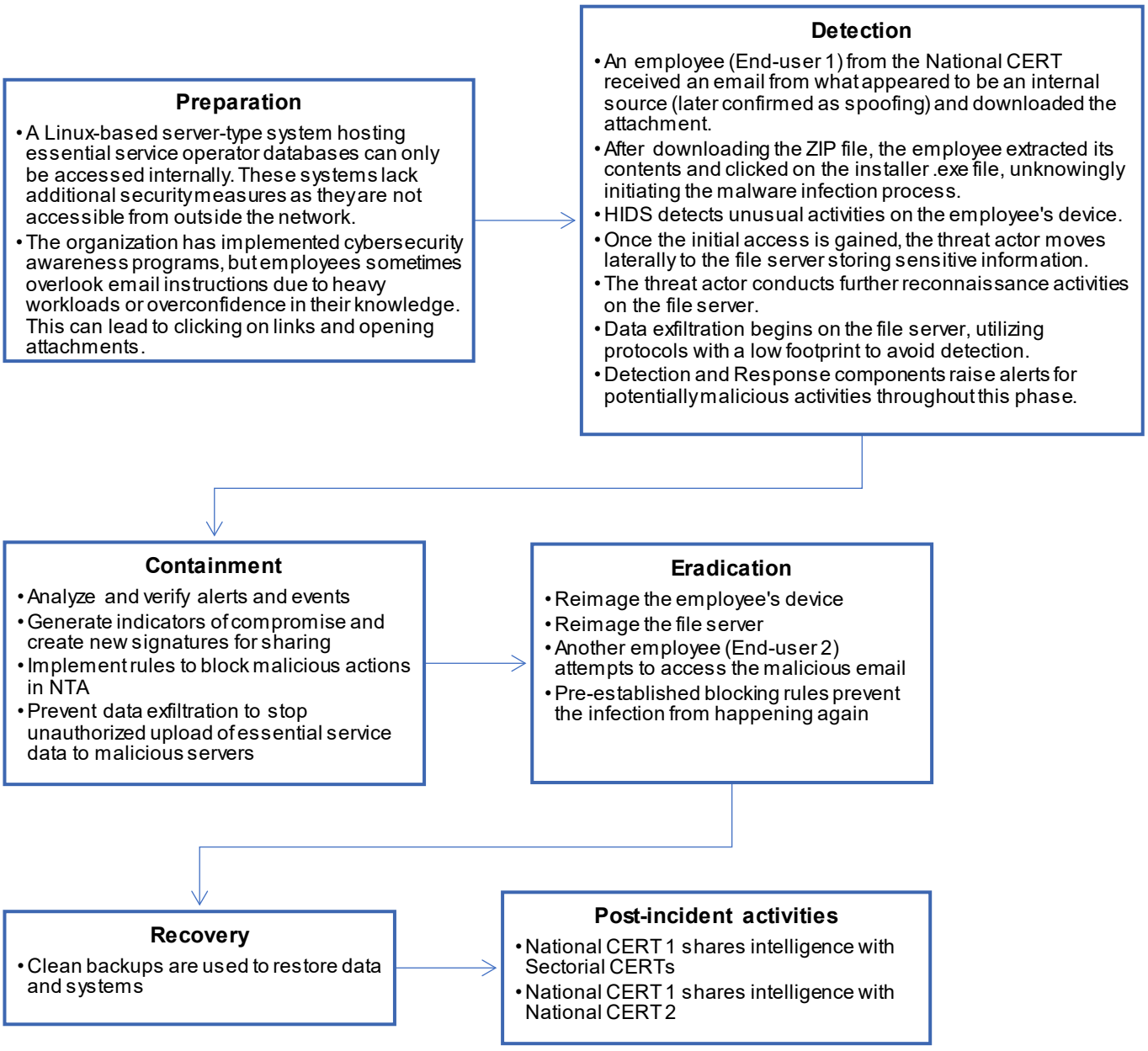

Figure 3 - Sequence of events mapped to incident response phases in Scenario 1 (Data Exfiltration on National CERT)

In Figure 4 below, we highlighted the sequence of events corresponding to each incident response phase in the second scenario (Preparation, Detection, Containment, Eradication, Recovery, and Post-Incident Activities).

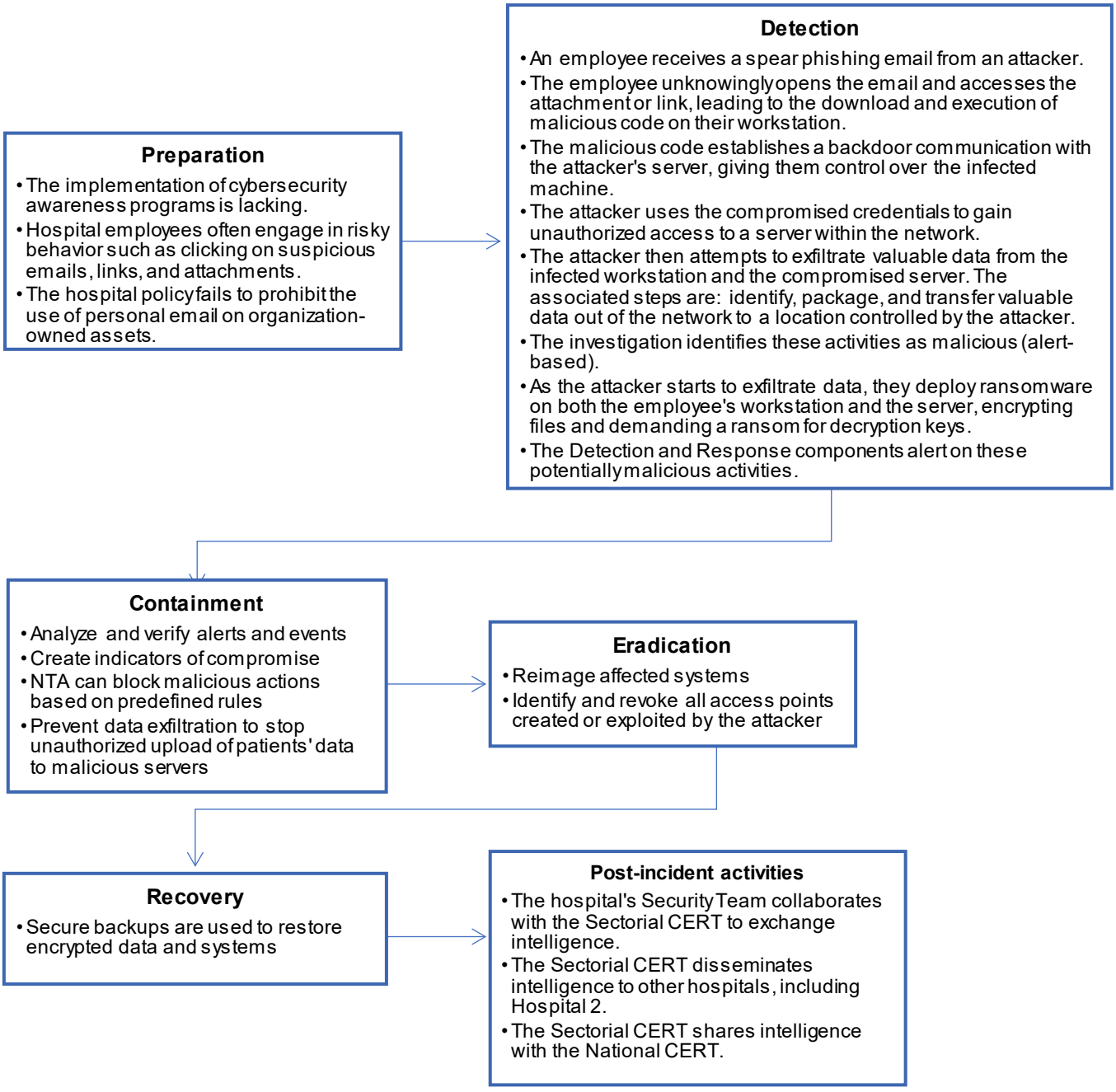

Figure 4 - Sequence of events mapped to incident response phases in Scenario 2 (Ransomware attack on public hospital)

Here are the services generated during the 2 pilot scenarios:

- Wartime intelligence
- Automated enrichment
- Local awareness

The specific services created during the pilot scenarios highlight the platform's ability to strengthen threat intelligence sharing, detection, analysis, and incident response processes.

Furthermore, the cyber threat intelligence and situational awareness services we developed allow us to showcase the platform's collaborative nature in terms of contractual KPIs, but more importantly in delivering practical benefits in real-world cyber-security contexts.

5. High-Level Architecture

5.1 CYDERCO Structure and Components

The following section outlines the alignment of development, technical work packages, and tasks with the solutions to be developed.

Figure 5 below illustrates the abstract described in the proposal document highlighting the platform's main components while introducing several new concepts such as endpoint data (E-data), respectively network data (N-data) and external sources for data ingestion (represented by External TI Providers in the following figure).

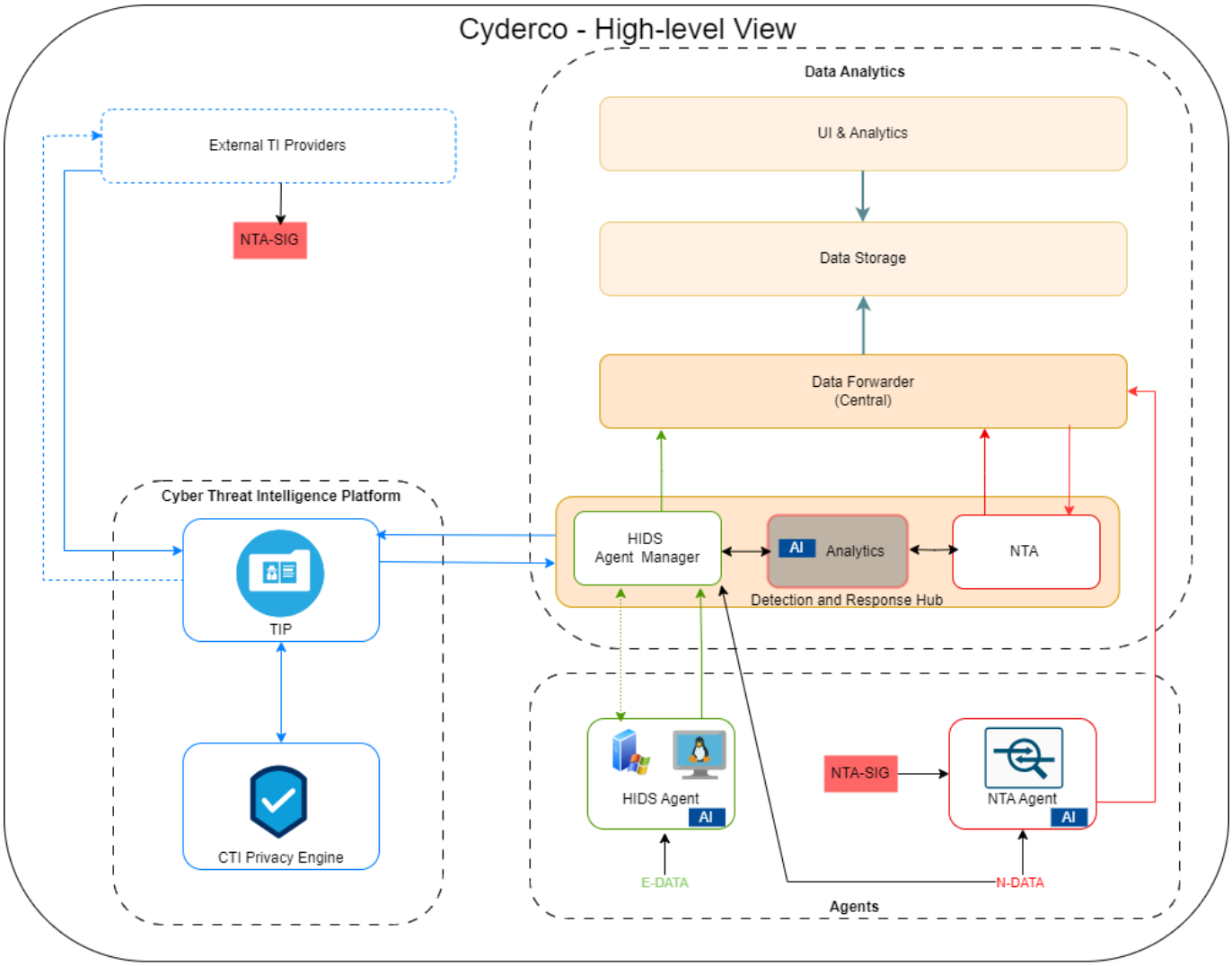

Figure 5 - High-level view on CYDERCO solution components

CYDERCO offers a consolidated platform composed of 2 main modules that respond to 2 different sets of challenges related to cybersecurity, as follows:

5.1.1 Detection and Response Hub

The **Detection and Response Hub** focuses on advanced detection and consists of the following main building blocks:

- **Data Analytics**

Data Analytics caters to Country SOCs and other relevant stakeholders during national-level cyber security incidents. Given the overwhelming amount of information received by a SOC, which is diverse and spans different timeframes, data exploration becomes essential to ensure an efficient and effective cyber detection capability.

The Data Analytics platform enables SOCs to detect, investigate, and respond to cyber security threats, including complex, targeted, or unknown attacks. It provides security teams with easy and intuitive access to the necessary information and relevant context, empowering them to identify all compromised assets and develop strategies to respond efficiently to threats and attacks.

The Data Analytics module will be perfected in task T3.1, under Work Package 3.

This solution will ingest and process or correlate vast amounts of data from various sources, including other modules within the Detection and Response Hub like HIDS, NTA, AI-Analytics, as well as the Threat Intelligence Platform and external sources. Additionally, it will transmit data such as alerts and IoCs to different platform components like AI-driven Analytics and TIP.

- **Network Traffic Analysis (NTA)**

Network Traffic Analysis or NTA is part of the Detection and Response Hub. It is conceived as a network-based detection and response solution that monitors network traffic and detects threats and malicious behavior.

To this effect, it leverages several techniques and capabilities, such as:

i. signature matching (The novelty, however, is that it can obtain newly developed signatures and enforce them during critical cyber incidents)
ii. anomaly-based detection
iii. Deep Packet Inspection including TLS, SSL, and HTTP (The novelty, however, is that it allows custom application layer parser development for protocols that are not supported by default)
iv. IDS and IPS capabilities

This module will be perfected in task T3.2, under Work Package 3. Besides detecting abnormal traffic patterns, NTA will generate alerts and forward them to the Data Analytics module (via Data the Forwarder), where SOC teams can consume the information.

NTA also communicates with the TIP module, which offers valuable insights such as IoC information that NTA uses to detect abnormal activities.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

- **Host Intrusion Detection System (HIDS)**

This module is part of the Detection and Response Hub and offers advanced detection capabilities to identify malicious activities at the host level. HIDS will collect, analyze, and correlate logs from various sources, including operating systems, applications, domains, incoming and outgoing network connections, and devices such as workstations and servers.

HIDS will use different techniques to detect threats, such as:

i. file integrity monitoring
ii. analysis and correlation of logs
iii. monitoring behavioral patterns in hardware resource consumption at host level
iv. monitoring behavioral patterns in network utilization

This module will be perfected in task T3.3, under Work Package 3. Like NTA, the HIDS module will forward alerts (with sufficient information) to the Data Analytics component, leveraging an intermediary component called Data Forwarder. It also communicates with the TIP component.

- **AI-driven Analytics**

The AI Analytics module corresponds to the Detection and Response Hub's fourth and final major component. It works closely with the Data Analytics module and is designed to send and receive information for the latter.

This module will be perfected in task T3.4, under Work Package 3. AI Analytics will provide superior detection and data processing capabilities and improve the efficiency of security operations.

To this effect, it leverages several techniques and capabilities, such as:

i. machine learning (ML)
ii. deep learning
iii. statistical anomalies

AI Analytics receives alerts from the Data Analytics component and preprocesses the data to prepare it for analysis performed by AI algorithms. AI then identifies relationships and dependencies between security events and intelligently correlates data obtained from different sources within the environment before sending it back to Data Analytics, where SOC and other security teams can consume it.

- Data Forwarder

The Data Forwarder serves as the central hub for data distribution among the various modules integrated within the platform. While depicted as a single entity, it is, in fact, a pervasive component that encompasses the underlying pipelines responsible for data transportation and transformation, operating under the Extract, Transform, Load (ETL) framework.

- **Data Storage* and *UI & Analytics*

Although these modules are not categorized as primary components, they are essential for efficiently storing data distributed across the various integrated modules of the platform.

They provide SOC teams with a user-friendly graphical interface that facilitates effective investigation through quick search and retrieval of information.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

5.1.2 Threat Intelligence Platform

The **Threat Intelligence Platform** focuses on actionable CTI sharing and consists of 4 building blocks:

Threat Intelligence Collection and Storing

TI Collection and Storing is part of the Threat Intelligence Platform (TIP). It will be perfected in task T4.1, which falls under Work Package 4. The main objective of this module is to collect intelligence from various sources within the platform as well as externally and store it at the entity level.

For instance, the TIP will ingest threat intelligence such as indicators of compromise from the Detection and Response Hub (Data Analytics, NTA, HIDS). As for external sources, the Collection and Storing module will support the ingestion of third-party threat intelligence feeds, such as OSINT, national SOCs, and other organizations that offer CTI.

- **Threat Intelligence Sharing**

TI Sharing is another important component of the TIP. It involves securely and instantly distributing threat intelligence to multiple subscribing entities while ensuring compatibility with different formats. The module will enhance the cybersecurity posture by providing high-confidence intelligence that can be quickly deployed in various systems.

To this effect, it leverages several techniques and capabilities, such as:

i. automatic synchronization of events across various instances
ii. interoperable file format for easy export
iii. privacy-preserving mechanisms such as field encryption and anonymization

This module will be perfected in task T4.2, under Work Package 4. TI Sharing will seamlessly communicate with other modules within the TIP, as well as other components of the CYDERCO platform such as Data Analytics, HIDS, and NTA, which will receive IoCs from TIS.

- **Threat Intelligence Enrichment**

TI Enrichment is the TIP part that focuses on automatic CTI data enrichment and contextualization to extract the highest possible value for SOCs, CSIRTs, and CERTs. The primary objective is to develop proactive strategies that can prevent potential attacks.

To this effect, it leverages several techniques and capabilities, such as:

i. AI/ML algorithms to correlate TI data and find relationships between various threat elements and attributes
ii. scoring shared CTI relevance according to a given asset type
iii. stating the confidence level of stored indicators

This module will be perfected in task T4.3, under Work Package 4. It will communicate with all the other components within the TIP to enhance the overall quality of threat intelligence.

- **Sighting Support**

Sightings represent observed indicators that provide valuable insights into attack trends, targeted organizations, and more. Sighting support is an integral part of the TIP, focused on monitoring the evolution of an indicator throughout the platform.

To this effect, the module leverages several techniques and capabilities, such as:

i. monitoring and alerting other entities when an indicator is detected on a system or network
ii. increase the sightings level and/or confidence level when an indicator is seen within the organization

This module will be perfected in task T4.4, under Work Package 4. It communicates with other modules within the TIP as well as with the Data Analytics module in the Detection and Response Hub. The latter supports data enrichment with threat intelligence from the TIP.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

6. Information System and Technology Architecture

6.1 Application Architecture

6.1.1 Interface Catalog

The table below outlines the catalog headers and offers guidance on the types of information that should be captured within the application communication and interaction flows.

Table 4 - Catalog outlining descriptive criteria for each interface between CYDERCO components

Header Name	Description
Unique identifier	Given name or title of each interface
Interface location	Name of location of interface, e.g.: URL, FQDN
Interface type	API, Socket, Dashboard, etc.
Data source	Name of the system providing data o the current interface.
Technical data source	Name of the source of data schema, e.g.: schema, table, file, etc.
Technical data destination	Name of the destination of data schema, e.g.: schema, table, file, etc.
Data flow type	Name of the data flow type, e.g.: 1 way sync, 2 ways sync, query API (with state type and if it's a web API, state REST, SOAP, or others) or custom
Frequency	Frequency of data exchanges, e.g.: real-time, daily, configurable interval, ad-hoc, etc.
Exchanged data format	Format of data exchanged: CSV, JSON, OAI, etc.
Data object types exchanged	Type of data exchanged, e.g.: alerts, events, etc.
Interface purpose	Description of the interface in business terms (if applicable): it's intended scope of usage
Data transformation	Method or application used to manipulate data, e.g.: PL-SQL, HIDS Manager, etc.
Method	Method of moving data from source to destination, e.g.: push or pull

SSL/TLS	Security protocol employed with YES/NO options

6.1.2 Application Interfaces

The interfaces between the various applications composing CYDERCO are responsible for the platform's overall operational effectiveness and making sure that all components work together efficiently.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Table 5 - Application interface fields within CYDERCO

Interface Description – CYDERCO Platform Overview													
Unique ID	**Int type**	**Int location**	**Data source**	**Technical data source**	**Technical data destination**	**Data flow type**	**Frequency**	**Exchanged data format**	**Data object type**	**Int purpose**	**Data transf**	**Method**	**SSL/TLS**
ETL1-HIDS	TCP-Socket	HIDS manager	HIDS agent	File	File	1-way sync	Real-time	JSON	event	Event sync from HIDS	HIDS agent	Push	Yes
ETL1-NTA	TAIL-S2S	Data Fwd agent	NTA	File	File	1-way batch	Custom interval	JSON	event	Event sync from NTA	Data FWD agent	Push	Yes
ETL2-HIDSm	TAIL-S2S	Data Fwd	HIDS manager	File	File	1-way batch	Custom interval	JSON	alert	Alert sync from HIDS manager	Data FWD	Push	Yes
ETL2-NTA	HTTPS-API	Data Fwd	NTA manager	File	File	1-way batch	Custom interval	JSON	Misc.	Data sync + transf (NTA agent to NTA)	Data FWD	Push	Yes
ETL2-NTAm	TAIL-S2S	HIDS manager	Data Fwd	File	File	1-way batch	Custom interval	JSON	Alert	Data sync (NTA manager to Storage via Data Fwd)	Data Fwd	Push	Yes
ETL-Storage	HTTPS-API	Data Storage	Data Fwd	File	Index	REST API	Custom interval	JSON	Alert	Long term storage of alerts	Data Fwd	Push	Yes
Analytics	HTTPS-API	Data Analytics	Data Fwd	Index	Web interface	REST API	Ad hoc	JSON	Alert	View & manage alerts	Data Analytics	Pull	Yes

TIP-OUT	HTTPS-API	EXT TIP	TIP	DB	DB	REST API	Real-time	JSON	IOC	Data sync (MISP to EXT providers)	Python lib/API	Push	Yes
TIP-IN	HTTPS-API	TIP	EXT TIP	DB	DB	REST API	Real-time	JSON	IOC	Data sync (EXT Providers to TIP)	Python lib/API	Pull	Yes
DR-TIP	HTTPS-API	TIP	HIDS manager	Index	DB	REST API	Real-time	JSON	IOC	Data sync (DR Hub to TIP)	Python lib/API	Push	Yes
TIP-HIDS	HTTPS-API	HIDS manager	TIP	DB	Index	REST API	Real-time	JSON	IOC	Data sync (TIP to DR Hub)	Python lib/API	Pull	Yes
TIP-PE	Zero-MQ	TIP	TIP	DB	DB	1-way sync	Real-time	JSON	IOC	Encrypt/Annonymize IOC	CTI Privacy Engine	Push	Yes
PE-TIP	HTTPS-API (PyMISP)	TIP	TIP	DB	DB	1-way sync	Real-time	JSON	IOC	Encrypt/Annonymize IOC	CTI Privacy Engine	Push	Yes
MGMT-HIDS	TCP-Socket	HIDS manager	HIDS manager	File	File	Encrypted socket	Real-time	XML	Config	HIDs config. and enrollment	N/A	Pull	Yes
E-DATA	Socket/Int	HIDS agent	OS Data	File/Socket, etc.	Socket	Local	Custom interval	Misc.	Event	Data collection from endpoints	N/A	Pull	No
NTA SIG	HTTPS-API	NTA agent	EXT Providers	File	Index	1-way sync	Ad hoc	Rules	Rules	Import signatures from EXT Providers	NTA	Pull	Yes

N-DATA	Socket/Int	NTA agent	Network	Network packets	File	Network captured data	Near real-time	Pcap	Network packets	Collect network traffic	NTA	Pull	No
AIA_HIDS	HTTPS-API	AIA	HIDS	File	File	1-way sync	Real-time	JSON	Alert	Identify relevant events on host	AIA	Push	Yes
AIA_NTA	HTTPS-API	AIA	NTA	File	File	REST API	Custom Interval	JSON	Alert	Send alerts for correlation	AIA	Push	Yes

The CYDERCO platform integrates multiple systems and components to deliver a cohesive detection and threat intelligence-sharing solution. A holistic approach is essential to ensure seamless collaboration among the various applications and enable real-time data collection from diverse sources.

We have designed the interfaces between these applications to eliminate silos, resulting in a comprehensive system that functions as a cohesive unit, regardless of the differing requirements and communication methods of its components.

Our approach effectively addresses these gaps in data and process integration. To facilitate data interfacing, we have implemented ETL (Extract, Transform, Load) pipelines.

Here are a few examples:

- ETL1-HIDS

This generic interface collects data and forwards it without any transformation.

It utilizes a TCP socket for a one-way, real-time synchronization flow, transferring events from the HIDS agent to the HIDS Manager.

- ETL2-HIDSm

This management interface connects the data source with the central events connector. It retrieves data from the HIDS manager and forwards it to the Central Data Forwarder.

The interface employs TAIL S2S for a one-way batch flow, synchronizing alerts from the HIDS Manager to the Central Data Forwarder at configurable intervals.

- MGMT-HIDS

Utilizing a TCP socket, this interface facilitates an encrypted real-time flow for HIDS configuration and enrollment.

While no actual data is exchanged, it enables bidirectional pipeline management between the agent and manager.

- ETL1-NTA

This generic interface collects data and forwards it without transformation. It employs TAIL S2S (a Linux-based command) for a one-way batch flow, synchronizing events from the NTA agent to the Central Data Forwarder at configurable intervals.

- ETL2-NTAm

This interface transfers information from the NTA Manager to the Central Data Forwarder. It also uses TAIL S2S for a one-way batch flow, synchronizing alerts from the NTA Manager to the Central Data Forwarder at configurable intervals.

- Finally, the Socket interfaces, N-DATA and E-DATA, are specifically designed to ingest raw data. N-DATA captures network traffic in the form of packets and transmits this information at the NTA Agent level. In contrast, E-DATA gathers operating system data from endpoints as events and relays it at the HIDS Agent level.

- Additionally, the TIP-PE interface serves not only as a transportation mechanism but also plays a crucial role in data transformation. It is engineered to encrypt and anonymize Indicators of Compromise (IOCs) exchanged between the Threat Intelligence Platform (TIP) and the Cyber Threat Intelligence (CTI) Privacy Engine.

To summarize, these pipelines allow us to consolidate data from multiple sources into a single, unified repository, enhancing our overall data management capabilities.

We have prioritized APIs as the preferred interfaces for process integration to streamline workflows across various applications and enhance operational management.

Here are a few examples:

- TIP-OUT

This process extracts Indicators of Compromise (IOCs) from the Threat Intelligence Platform and transmits them to external Threat Intelligence providers via an HTTPS REST API.

The data exchange occurs in real time, with connections secured using SSL/TLS protocols.

- TIP-IN

This process imports IOCs from external Threat Intelligence providers into the Threat Intelligence Platform through an HTTPS REST API.

Like TIP-OUT, the data exchange is conducted in real-time and secured via SSL/TLS.

- DR-TIP

This interface pulls data from the HIDS Manager in the Detection and Response Hub and conveys it to the Threat Intelligence Platform using an HTTPS REST API.

Data is exchanged in real time through a JSON file format.

- TIP-HIDS

This interface transfers data from the Threat Intelligence Platform to the HIDS Manager in the Detection and Response Hub via an HTTPS REST API.

The exchange is conducted in real time using a JSON file.

- ANALYTICS

This interface leverages an HTTPS REST API to facilitate alert management and visualization.

It specifically handles data objects related to alerts that have been collected from the Central Data Forwarder, subsequently stored in a Data Storage index, and made available for visualization through the graphical interface (Data Analytics) as required.

- ETL STORAGE

Similar to the Analytics process, this interface employs an HTTPS REST API for the long-term storage of alerts gathered from the Central Data Forwarder, facilitating their transport to a Data Storage index.

- ETL2-NTA

This interface transfers information from the Central Data Forwarder to the NTA Manager using HTTPS API.

These are just a few examples of the interfaces developed to facilitate the integrations that collectively enhance our operational efficiency and data management capabilities.

- NTA SIG

This interface facilitates the collection of signatures from External TI Providers and integrates the data at the NTA agent level through a secure HTTPS API.

The data exchange occurs on an ad-hoc basis and is initiated on demand using a .rules file, with synchronization executed in a unidirectional manner.

*Note: While the graphical representation (*see Figure 7 below*) illustrates two instances of NTA SIG, in the context of the platform there is a single pipeline specifically designed for the transfer of rules.

- PE-TIP

Similar to the TIP-PE interface, this pipeline serves not only as a transportation mechanism but also plays a crucial role in data transformation.

It is engineered to encrypt and anonymize Indicators of Compromise (IOCs) exchanged between the Cyber Threat Intelligence (CTI) Privacy Engine and Threat Intelligence Platform (TIP).

This interface transfers information using a customized HTTPS API (PyMISP).

- AIA_HIDS

This interface leverages an HTTPS API to facilitate transportation of relevant events identified on the host in the form of alerts.

It specifically handles data exchanged in JSON files from the HIDS manager to the AI Engine within the Detection and Response Hub.

- AIA_NTA

Finally, this interface acts as transportation mechanism for data object types related to alerts. It facilitates transportation between the NTA manager and the AI engine, preparing data for alert correlation and consolidation.

The format of the data exchange in also JSON and the frequency of exchanges can be customized according to the needs of the stakeholders.

This interface leverages an HTTPS REST API.

6.1.3 Application Communication Diagram

The diagram below illustrates the interfaces aligned with the graphical representation of the catalog fields outlined in section 6.1.1.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

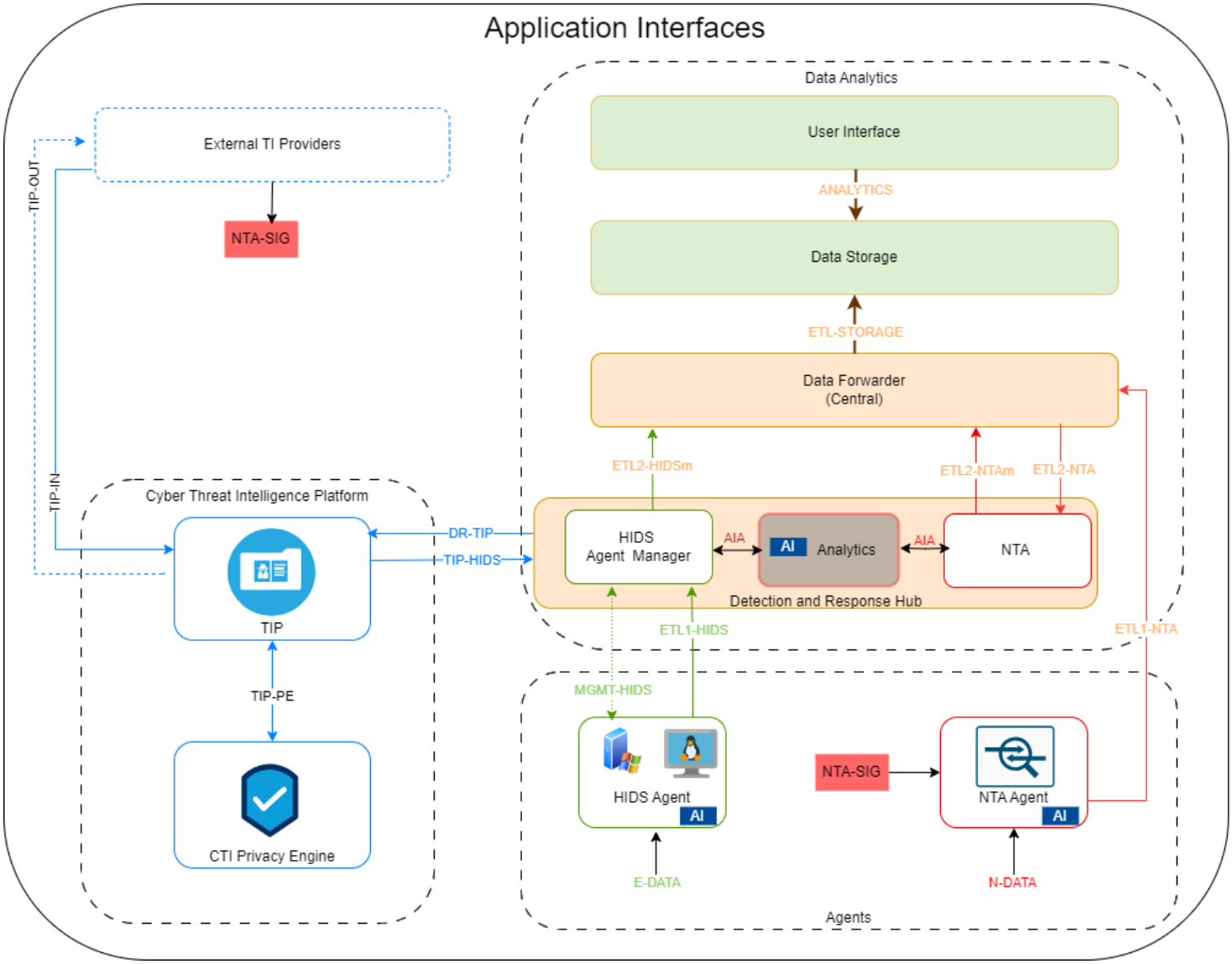

Figure 6 - Application Communication Diagram (CYDERCO Interfaces)

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

6.2 Data Architecture

6.2.1 Fair Data Principles

To align with contemporary standards in scientific data management, we are committed to integrating the FAIR principles (GO FAIR, 2024) of findability, accessibility, interoperability, and reusability through the technologies employed within the CYDERCO Platform.

This approach will optimize the analysis of diverse datasets, enhance machine capabilities for data discovery and utilization, and promote knowledge and innovation through effective data sharing and reuse.

The table below illustrates these principles, as follows:

Table 6 – FAIR data principles implemented in CYDERCO and their associated description

Category	FAIR Principle	Description/Purpose
Findable	(Meta)data are assigned a globally unique and persistent identifier.	The purpose of this principle is to eliminate any ambiguity within the dataset.
	Data are described with rich metadata.	This principle aims to increase data findability through extensive associated metadata describing the quality and characteristics of the respective data.
	Metadata clearly and explicitly include the identifier of the data they describe.	This principle reinforces the necessity to associate each dataset with its corresponding identifier that is globally unique and persistent.
	(Meta)data are registered or indexed in a searchable resource.	This principle states the importance of dataset indexation to ensure its discovery in the digital environment.
Accessible	(Meta)data are retrievable by their identifier using a standardized communications protocol.	This principle states that data retrieval must be performed using non-proprietary communication methods (such as HTTP(S) or FTP).
	The protocol is open, free, and universally implementable.	This principle reinforces the importance of using globally implementable protocols to facilitate data retrieval and reuse.

	The protocol allows for an authentication and authorization procedure, where necessary.	This principle aims to unequivocally outline the conditions under which specific data can be accessed.
Interoperable	(Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	This principle establishes a shared framework for knowledge representation through a language characterized by clear syntax rules and readily accessible specifications for all users (humans or machines).
	(Meta)data include qualified references to other (meta)data.	This principle seeks to enhance knowledge by providing context for metadata. It stipulates that any dataset must indicate, with appropriate citations, if it is derived from or relies on information from another dataset.
Reusable	(Meta)data are richly described with a plurality of accurate and relevant attributes.	This principle aims to enhance findability by encouraging extensive data labeling. It extends a previous principle stating the importance of discovery-enabling metadata and introduces context metadata to describe the data generated accurately.
	(Meta)data are associated with detailed provenance.	This principle encourages data reuse by clearly stipulating the source and the conditions under which the published dataset has been generated.

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

6.2.2 Data Catalog

The table below outlines the catalog headers and offers guidance on the data field types that should be generated and captured by the applications engaged in the data flows.

Table 7 - Catalog outlining descriptive criteria for each data field within CYDERCO components

Header name	Description
Field source	Component which created the field
Field name	Represents the name of the field, e.g.: SRC_IP, Process Name, etc.
Field description	Explains the field, e.g.: SRC_IP captures the source IP for the alert/event of interest
Field type	States the type of the field value, e.g.: integer, string, timestamp, etc.
Importance	States the criticality of the field: High-cannot achieve goal without field, Medium – can acheive goal but it is desirable for quality improvement, Low – does not affect goal or quality
Status	States if the field is new or existing
Field value sample	Represents a sample of the value contained by the field, e.g.: for IP address-192.168.100.1

6.2.3 Data Fields mapped to components

To effectively demonstrate the data fields generated and captured throughout our data flows, we have utilized the following components:

- Standard fields presented in a JSON format file (NTA-based, HIDS-based)
- Proprietary JSON sample files produced within the CYDERCO pilot environment

Our methodology involves a comparative analysis of these two components, filtering the information through the lens of cybersecurity analysts. This approach enables us to curate a tailored set of data fields that are essential for advancing investigations and incident response processes.

For illustration purposes, we have focused on 2 major CYDERCO components for the time being: NTA and HIDS. The data flows we chose to highlight in the initial version of the current deliverable D2.3 are based on the alert module as far as the NTA component is concerned, respectively the FIM module for the HIDS component.

For the final version of deliverable D2.3, we will extend the use cases of each of these two components and further list the data fields associated with the remainder of CYDERCO components to depict a comprehensive landscape of data flows at the platform level.

- **NTA**

The NTA event sample file is attached below for further reference:

NTA event sample file.pdf

The following block outlines the standard fields presented in a JSON format event file (NTA-based):


```
{
  "timestamp": "2017-04-07T22:24:37.251547+0100",
  "flow_id": 586497171462735,
  "pcap_cnt": 53381,
  "event_type": "alert",
  "src_ip": "192.168.2.14",
  "src_port": 50096,
  "dest_ip": "209.53.113.5",
  "dest_port": 80,
  "proto": "TCP",
  "metadata": {
    "flowbits": [
      "http.dottedquadhost"
    ]
  },
  "tx_id": 4,
  "alert": {
    "action": "allowed",
    "gid": 1,
    "signature_id": 2018358,
    "rev": 10,
    "signature": "ET HUNTING GENERIC SUSPICIOUS POST to Dotted Quad with Fake Browser 1",
    "category": "Potentially Bad Traffic",
    "severity": 2
  },
  "app_proto": "http"
}
```

To ensure consistency, we have outlined the data fields in accordance with the specific criteria detailed in the Data Catalog, as follows:

- Field Source
- Field Name
- Field Description
- Field Type
- Importance
- Status
- Field value sample

The below table presents the data fields generated by the NTA component, in the context of an alert module (corresponding pipeline or interface for communication is ETL1-NTA).

Table 8 – Alert module-based data fields within NTA

NTA						
Field Source	**Field Name**	**Field description**	**Field type**	**Importance**	**Status**	**Field Value Sample**
src_ip	Source IP	Source IP of the network packets	String	High	Existing	192.168.2.14
dest_ip	Destination IP	Destination IP of the network packets	String	High	Existing	209.53.113.5
src_port	Source port	Source port of the network packets	Integer	High	Existing	50096
dest_port	Destination port	Destination port of the network packets	Integer	High	Existing	80
alert.signature	Signature name	Detection type signature for signatures already in NTA	String	High	Existing	ET HUNTING GENERIC SUSPICIOUS POST
proto	Protocol	Network protocol for	String	High	Existing	TCP

		which the alert was generated				
alert.action	Action	Action taken for the network packet(s): alert, block, drop	String	High	Existing	allowed
metadata	Metadata	Creation, author, organization, etc.	Array	Medium	Existing	
alert.signature_id	Signature ID	Helps with correlation	Integer	High	Existing	2018358
alert.severity	Severity	Indicates the level of severity	Integer	High	Existing	2
alert.category	Category	Helps with pivoting	String	Medium	Existing	Potentially Bad Traffic
timestamp	Time Stamp	The date and time when the event was detected	Time Stamp	High	Existing	2017-04-07T22:24:37.251547+0100
flow_id	Flow ID	Unique identifier for a network flow used to associate packets and events with a specific connection	String	High	Existing	586497171462735
app_proto	App protocol	The application layer protocol detected in the network traffic: HTTP, DNS, FTP, LDAP, etc.	Integer	Medium	Existing	http

- **HIDS**

The HIDS event sample file is attached below for further reference:

HIDS event sample file.pdf

The following bloc outlines the fields presented in a JSON format file (HIDS-based):


```
{
 "syscheck": {
  "path": "c:\\temp\\test1.txt",
  "diff": "---\n> \n> asdgsdfgsdh\n> dfgh\n> \n> fgdh\n> \n",
  "event": "modified",
  "mode": "realtime",
  "uname_after": "admin",
  "mtime_after": "2024-09-16T13:09:52",
  "uid_after": "S-1-5-21-2419750563-2712209062-3083259898-1001",
  "win_perm_after": [
   {
    "name": "Administrators"
   },
   {
    "name": "SYSTEM"
   },
   {
    "name": "Users"
   },
   {
    "name": "Authenticated Users"
   }
  ],
  "md5_before": "c6e636097237ea9b9498e418af11958d",
  "md5_after": "dc3723296bece9d579f089619455b609",
  "sha1_before": "64afc1b29cdd92f8e452c0e0fe1768f7e0f093c3",
  "sha1_after": "1abd8afe33ee40a66b12fea687afee7511c7c6c9",
  "sha256_before": "6075fbf60a7c5050a243b20a3976484947016 60d662fbe849d2e6f4cd52ef2e4",
```

```
"sha256_after": "3b2eac55c8492e7785138f571b7712f540f11dd7e6ba4e0021db5e1839795ca7"
  },
  "timestamp": "2024-09-16T11:09:52.837+0000",
  "_index": "wazuh-alerts-4.x-2024.09.16",
  "rule": {
    "groups": [
      "ossec",
      "syscheck",
      "syscheck_entry_modified",
      "syscheck_file"
    ]
  },
  "id": "1726484992.4144427",
  "rule": {
    "level": 7,
    "description": "Integrity checksum changed.",
    "id": "550",
    "firedtimes": 1,
    "mail": false
  },
  "agent": {
    "id": "005",
    "ip": "172.16.110.190",
    "name": "cy-w10-01"
  },
  "input": {
    "type": "log"
  },
  "location": "syscheck",
  "manager": {
    "name": "wazuh.manager"
  },
```

```
"full_log": "File 'c:\\temp\\test1.txt' modified\nMode: realtime\nChanged attributes: size,mtime,md5,sha1,sha256\nSize changed from '24' to '57'\nOld modification time was: '1726234191', now it is '1726492192'\nOld md5sum was:


'c6e636097237ea9b9498e418af11958d'\nNew md5sum is : 'dc3723296bece9d579f089619455b609'\nOld sha1sum was: '64afc1b29cdd92f8e452c0e0fe1768f7e0f093c3'\nNew sha1sum is : '1abd8afe33ee40a66b12fea687afee7511c7c6c9'\nOld sha256sum was: '6075fbf60a7c5050a243b20a397648494701660d662fbe849d2e6f4cd52ef2e4'\nNew sha256sum is : '3b2eac55c8492e7785138f571b7712f540f11dd7e6ba4e0021db5e1839795ca7'\n",
 "decoder": {
  "name": "syscheck_integrity_changed"
 }
}
```


To ensure consistency, we have outlined the data fields in accordance with the specific criteria detailed in the Data Catalog, as follows:

- Field Source
- Field Name
- Field Description
- Field Type
- Importance
- Status
- Field value sample

The below table presents the data fields generated by the HIDS component, in the context of a FIM module (corresponding pipelines or interfaces for communication are ETL1-HIDS and ETL2-HIDSm).

The project funded under Grant Agreement No. 101128052 is supported by the European Cybersecurity Competence Centre

Table 9 - FIM module-based data fields within HIDS

HIDS						
Field source	**Field name**	**Field description**	**Field type**	**Importance**	**Status**	**Field Value Sample**
syscheck.(md5|SHA1|SHA256)_before	File hash	File hash value before the event (values depend on hash type)	string	high	existing	c6e636097237ea9b9498e418af11958d/ 64afc1b29cdd92f8e452c0e0fe1768f7e0f093c3/ 6075fbf60a7c5050a243b20a397648494701660d662fbe849d2e6f4cd52ef2e4
syscheck.(md5|SHA1|SHA256)_after	File hash	File hash value after the event (values depend on hash type)	string	high	existing	dc3723296bece9d579f089619455b609/ 64afc1b29cdd92f8e452c0e0fe1768f7e0f093c3/ 3b2eac55c8492e7785138f571b7712f540f11dd7e6ba4e0021db5e1839795ca7
syscheck.path	File path	Specific local path of the file or registry keys	string	high	existing	c:\temp\test1.txt
syscheck.diff	File difference	Indicates the difference between compared file versions (before, after)	string	high	existing	asdgsdfgsdh dfgh fgdh
syscheck.event	Event type	Specific type of event: modified, deleted, etc.	string	high	existing	modified
syscheck.mode	Detection mode	Detection mode can be real-time or custom	string	high	existing	real-time
syscheck.uname_after	Username after	File owner after modification (on file)	string	high	existing	admin
syscheck.mtime_after	Mtime after	Time of modification on file (system time)	timestamp	high	existing	2024-09-16T13:09:52

syscheck.uid_after	UID	Unique identifier for user	string	high	existing	S-1-5-21-2419750563-2712209062-3083259898-1001
syscheck.win_perm_after.name	File permission	File permissions after event	string	high	existing	Administrators (Permission list)/ SYSTEM (Permission list)/ Users (Permission list)/ Authenticated users (Permission list)
timestamp	Timestamp	Execution timestamp	timestamp	high	existing	2024-09-16T11:09:52.837Z
_index	Index	Open-search index that stores the event	string	high	existing	wazuh-alerts-4.x-2024.09.16
rule.groups	Rule category	Indicates rule classification	string	high	existing	ossec/syscheck/syscheck_entry_modified/syscheck_file
id	ID	Unique identifier for event	integer	high	existing	1726484992
rule.level	Severity	Threat level (1-15)	integer	high	existing	7
agent.id	Agent ID	ID of HIDS agent that generated the event	string	high	existing	005
agent.ip	Agent IP	IP address of HIDS agent that generated the event	string	high	existing	172.16.110.190
agent.name	Agent name	Name of HIDS agent that generated event	string	high	existing	cy-w10-01
input.type	Input type	Type of data received by the HIDS manager	string	high	existing	log
location	Location	Source path of data received by the HIDS manager	string	high	existing	syscheck

manager.name	Manager name	Name of HIDS manager that generated the event	string	high	existing	Wazuh.manager
rule.description	Rule description	Description of HIDS manager rule	string	high	existing	Integrity checksum changed
rule.id	Rule ID	ID of HIDS manager rule	integer	high	existing	550
rule.firedtimes	Fired times	Number of times the rule was triggered (same rule, same data)	integer	high	existing	1
rule.mail	Rule mail	Indicates if HIDS manager should send an email for rule or not	boolean	high	existing	FALSE
full_log	Full log	Initial message as sent by HIDS agent - option based on rule definition	string	high	existing	File 'c:\\temp\\test1.txt' modified\nMode: realtime\nChanged attributes: size,mtime,md5,sha1,sha256\nSize changed from '24' to '57'\nOld modification time was: '1726234191', now it is '1726492192'\nOld md5sum was: 'c6e636097237ea9b9498e418af11958d'\nNew md5sum is : 'dc3723296bece9d579f089619455b609'\nOld sha1sum was: '64afc1b29cdd92f8e452c0e0fe1768f7e0f093c3'\nNew sha1sum is : '1abd8afe33ee40a66b12fea687afee7511c7c6c9'\nOld sha256sum was: '6075fbf60a7c5050a243b20a397648494701660d662fbe849d2e6f4cd52ef2e4'\nNew sha256sum is : '3b2eac55c8492e7785138f571b7712f540f11dd7e6ba4e0021db5e1839795ca7'\n

decoder.name	Decoder name	HIDS decoder name	string	high	existing	syscheck_integrity_changed

6.2.4 Dataflow Diagram

The below diagram shows the data fields generated by the NTA component, in the context of an alert module.

The image highlights the corresponding pipelines or interfaces for communication: **ETL1-NTA** and **ETL2-NTA** while the remainder of components is faded out to deliberately narrow the perspective and increase the focus on selected items.

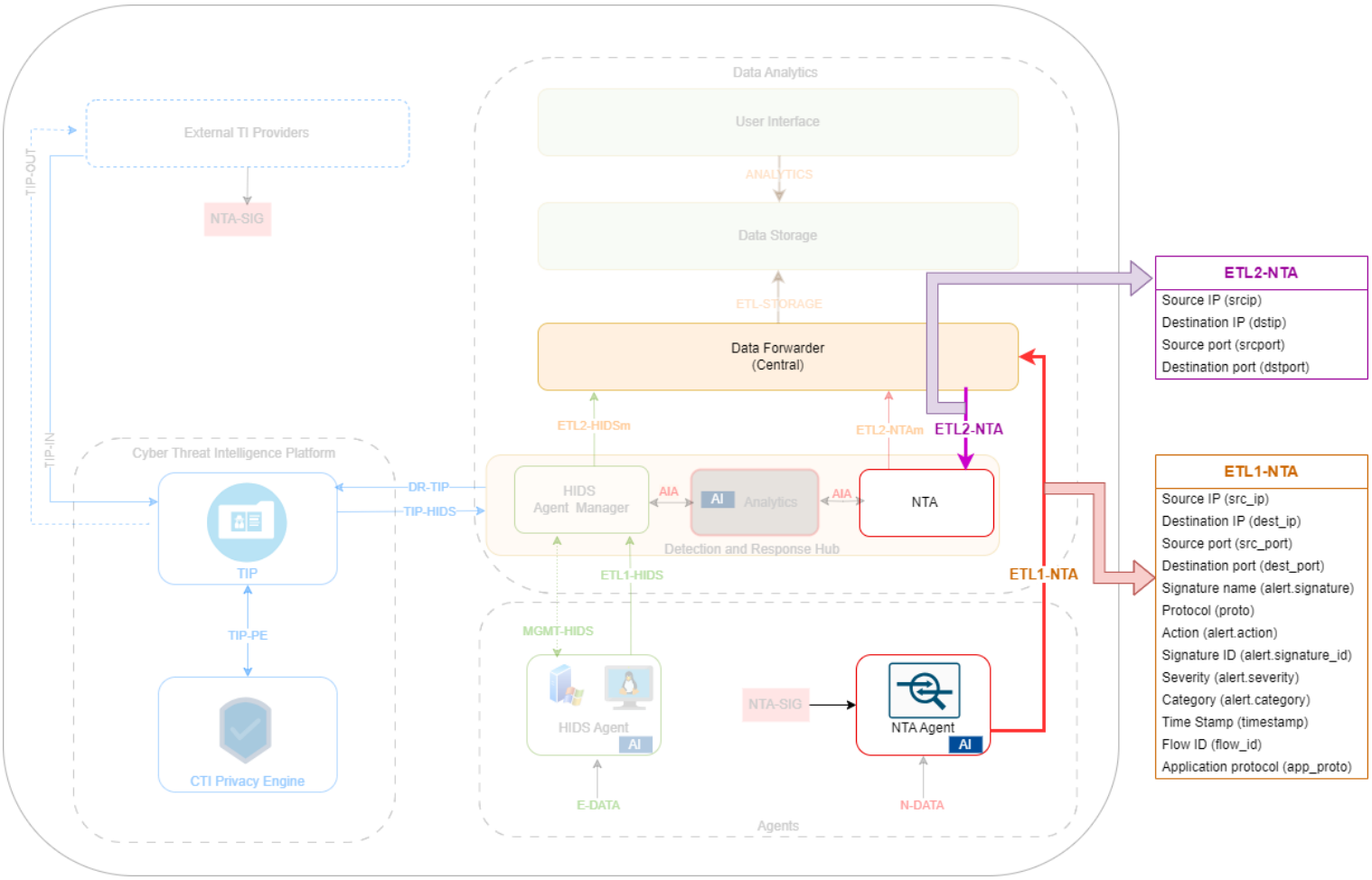

Figure 7 - NTA data communication flows and pipelines

The matrices presented in the diagram only list the Field Source and Field Name data field types, while the corresponding table in the document (***table 8***) provides a comprehensive view of the NTA component through the data architecture's lens.

The table extensively explains all the field enumerated in our data catalog: *Field Source, Field Name, Field Description, Field Type, Importance, Status,* and *Field value sample*.

The below diagram shows the data fields generated by the HIDS component, in the context of a FIM module.

The image highlights the corresponding pipelines or interfaces for communication: **ETL1-HIDS** and **ETL2-HIDSm** while the remainder of components is faded out to deliberately narrow the perspective and increase the focus on selected items.

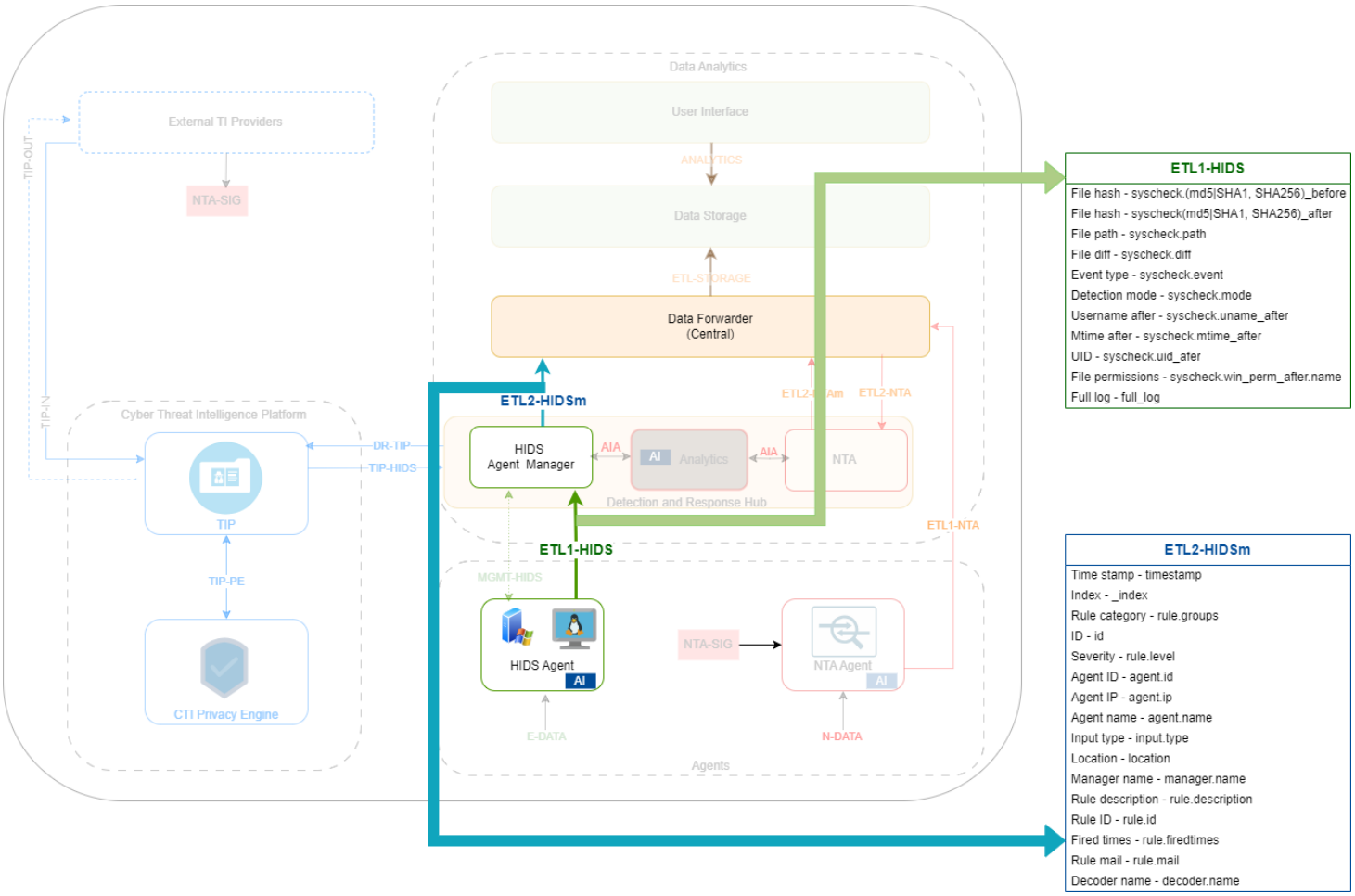

Figure 8 - HIDS data communication flows and pipelines

The matrices presented in the diagram only list the Field Source and Field Name data field types, while the corresponding table in the document (***table 9***) provides a comprehensive view of the HIDS component through the data architecture's lens.

The table extensively explains all the field enumerated in our data catalog: *Field Source, Field Name, Field Description, Field Type, Importance, Status,* and *Field value sample*.

6.3 Technology Architecture

6.3.1 Technology Architecture Components and Diagram

The diagram below (*figure 9*) provides a visual representation of the technical architecture that underpins the CYDERCO platform. It illustrates all the interconnected components that form the foundation of our platform, including the supporting hardware, virtualization layer, and microservices.

These microservices include key project elements such as HIDS, NTA, TIP, CTI Privacy Engine, Data Analytics, and AI Analytics and additional supporting services like SaaS, PKI, and Identity management.

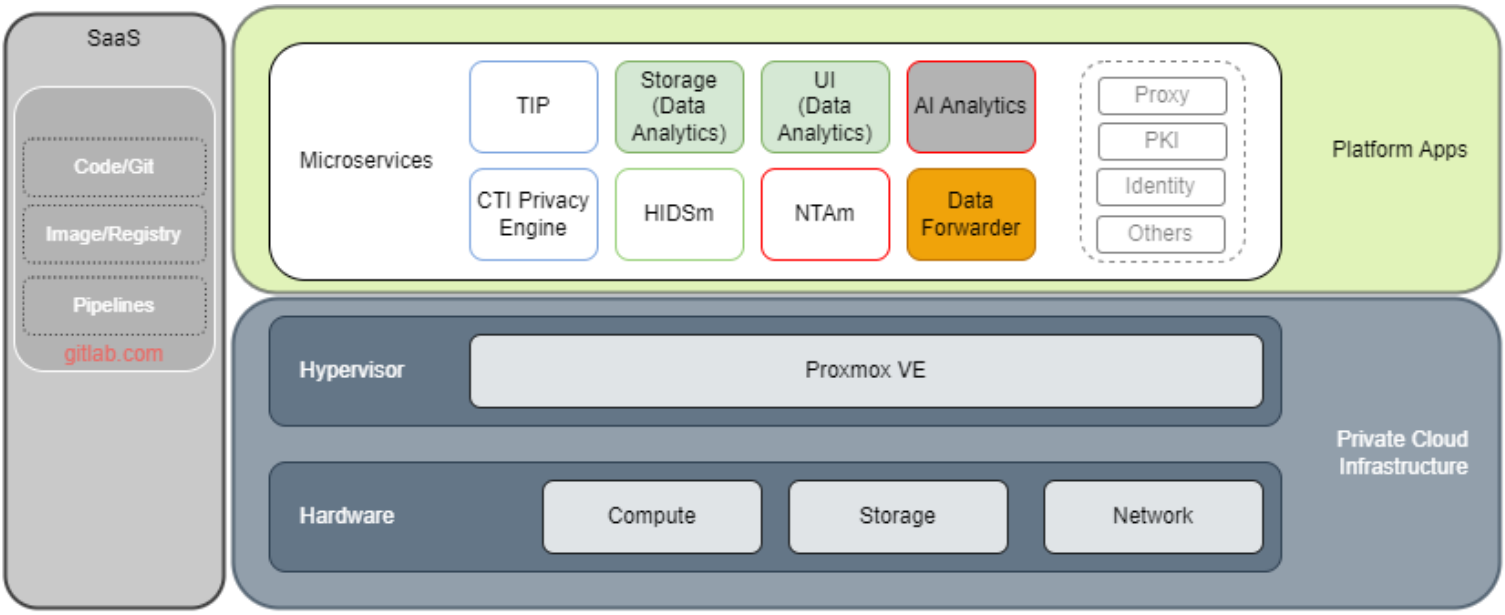

Figure 9 - CYDERCO technical architecture components

Components:

- **Private Cloud Infrastructure**

This layer is responsible for delivering essential hardware resources, including computing power, storage, networking, and the foundational virtualization layer.

Table 10 - Computing power (based on X86-84 architecture)

CYDERCO Component	(v)CPU - Cores	Memory (GB)	Deployment Model	Comments
HIDS Manager	1	2	Docker Container	
NTA Manager	2	4	Docker Container	
Data Forwarder (Central)	4	16	Docker Container	
Data Forwarder (Edge)	2	2	Docker Container	Can be deployed as daemon
Data Storage	4	16	Docker Container	
User Interface	2	4	Docker Container	
AI Analytics Engine	N/A	N/A	Docker Container	Not available at this moment
TIP	4	16	Docker Container	
CTI Privacy Engine	4	16	Docker Container	

**Note*: this section refers to physical storage requirements and not the CYDERCO Data Storage building block, unless specified otherwise.

Table 11 - Storage

CYDERCO Component	Size (GB)	Performance	Volume	Comments
HIDS Manager	32	X		
NTA Manager	32	X		
Data Forwarder (Central)	64	X	X	
Data Forwarder (Edge)	8	X		
Data Storage	100		X	

User Interface	N/A			
AI Analytics Engine	N/A			
TIP	64	X	X	
CTI Privacy Engine	N/A			

**Note*: The values listed below represent the minimum requirements for research and development purposes. Specific production requirements will be outlined in the final version of the platform design documentation.

- Network (*Figure 10 - CYDERCO network architecture components*)
- Hypervisor

We have selected Proxmox VE as our hypervisor. This open-source KVM/LXD-based (para)virtualization platform offers an efficient and scalable solution for the CYDERCO private cloud. Our deployment strategy will utilize both virtual machines and Linux containers to manage our Docker-based applications.

Some key features of Proxmox include granular control over resource allocation, seamless workload migration between nodes (hardware servers) and comprehensive support for a wide range of hardware and software components.

Currently, the CYDERCO Private Cloud Infrastructure it is designed for research and development purposes or to meet the needs of small to medium-sized businesses.

We recommend making adjustments following the "Scalability and Performance" guidelines when transitioning to production environments.

Organizations may utilize any Infrastructure as a Service (IaaS) solution from a public cloud provider as a substitute to enhance flexibility. This option will have no significant impact on other layers of the platform.

➢ **Docker CE**

The Platform Applications layer encompasses all necessary applications for the CYDERCO Platform. CYDERCO components are delivered as microservices in the form of Docker images, streamlining daily operations.

Docker Community Edition facilitates the development runtime environment.

For deployment automation, we rely on *Docker Compose* capabilities to centralize and simplify configuration management in a central location, while still allowing for the extension of configurations through additional files.

➢ **External Services (SaaS)**

- Git Repositories - These repositories are utilized to manage the code necessary for both infrastructure as code and application development.
- Image Registry - This service is dedicated to storing all custom images developed by CYDERCO.

- Pipelines - Our pipelines facilitate the automation needed for the seamless building of images stored in the image registry.

In our implementation, these external services are delivered through the GitLab.com (GitLab B.V., 2024) SaaS platform.

➢ **Security Considerations**

- Identity Management

Identity management is crucial for securing cloud infrastructures and the applications that operate within them. For the CYDERCO platform, we will implement robust security measures through external integrations, with a focus on the following key areas:

- Single Sign-On (SSO) with Multi-Factor Authentication (MFA)
- Digital Certificates

- Data Protection

Our data protection strategy encompasses cybersecurity best practices designed to secure data in transit across various data collection pipelines.

*Note: Security measures for "data at rest" and "data in use" fall outside the current scope and will be addressed during the project industrialization phase.

➢ **Scalability and Performance Considerations**

As the project progresses, we will implement specific considerations for production environments tailored to each platform component.

The following foundational elements will be addressed to ensure optimal scalability and performance:

- HIDS Manager
- Data Storage
- Data Forwarder
- Threat Intelligence Platform
- AI Analytics Engine

6.3.2 Network Architecture

The diagram below offers a visual overview of the network architecture that supports the CYDERCO platform.

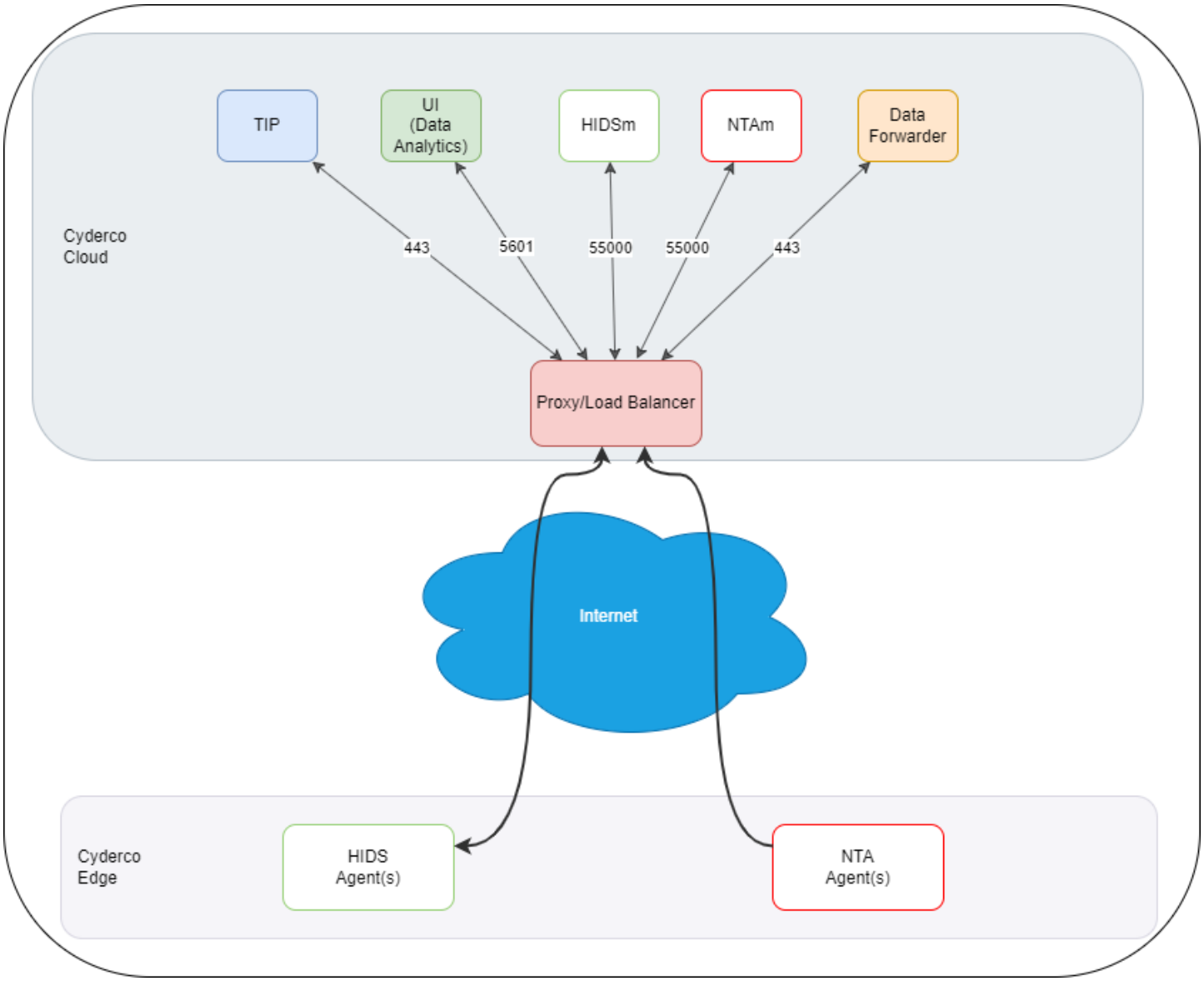

Figure 10 - CYDERCO network architecture components

It details interconnected components, including the CYDERCO Cloud, which houses essential microservices such as the NTA and HIDS managers, Data Forwarder, Data Analytics, and TIP.

The diagram also highlights their primary ports and the Proxy, which functions as a load balancer. Additionally, the Edge layer is depicted, featuring the HIDS and NTA agents responsible for general data collection, which are designed to be Internet-facing.

7. References

ENISA. (2024). *ENISA - European Union Agency for Cybersecurity*. Retrieved from ENISA: https://www.enisa.europa.eu/

European Insurance and Occupational Pensions Agency, DORA. (2024). *Digital Operational Resilience Act (DORA)*. Retrieved from EIOPA: https://www.eiopa.europa.eu/digital-operational-resilience-act-dora_en

GitLab B.V. (2024, Ocotober). *About GitLab*. Retrieved from GitLab Enterprise: https://about.gitlab.com/

GO FAIR. (2024). *Fair Principles*. Retrieved from FAIR Principles: https://www.go-fair.org/fair-principles/