D4.2

Design and Architecture for the Enterprise Pilot

WP4 – Enterprise Pilot

C3ISP

Collaborative and Confidential Information Sharing and Analysis for Cyber Protection

Due date of deliverable: 30/09/2017
Actual submission date: 30/09/2017

30/09/2017

Version 1.0

Responsible partner: SAP
Editor: Francesco Di Cerbo
E-mail address: francesco.di.cerbo@sap.com

Project co-funded by the European Commission within the Horizon 2020 Framework Programme		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

The C3ISP Project is supported by funding under the Horizon 2020 Framework Program of the European Commission DS 2015-1, GA #700294

Authors: P. Kearney, X. Wang, I. Herwono (BT), F. Di Cerbo (SAP)

Approved by: R. De Lemos (UniKENT), I. Matteucci (CNR)

Revision History

Version	Date	Name	Partner	Sections Affected / Comments
0.1	2/09/2017	Paul J. Kearney	BT	Initial draft
0.2	5/09/2017	Francesco Di Cerbo	SAP	Various contributions
0.3	16/09/2017	Paul J. Kearney	BT	System Overview, System Architecture
0.4	20/09/2017	Ian Herwono	BT	Component Architecture, Data Model
0.5	20/09/2017	Xiao-Si Selina Wang	BT	Future Work
0.6	24/09/2017	Francesco Di Cerbo	SAP	Introduction, Security Model, Requirements Matrix, Integration Plan, Conclusions
0.7	26/09/2017	Rogerio De Lemos	UniKent	Review
0.8	27/09/2017	Ilaria Matteucci	CNR	Review
0.9	28/09/2017	X. Wang, I. Herwono, F. Di Cerbo	BT, SAP	Changes following to comments
1.0	29/09/2017	F. Di Cerbo	SAP	Final version

Executive Summary

The document presents the first version of the Enterprise pilot architecture. One of the main challenges addressed by the document is to introduce the new advanced features of the C3ISP framework and to integrate them within the software infrastructure of a security provider of Managed Security Services currently available on the market.

The initial architecture achieves this main objective. The relevant Use Cases presented in Deliverable D4.1 are addressed by new architecture diagrams that will be used to guide the implementation phase.

Data model, security model as well as run-time requirements are also presented in the document. Moreover, special attention has been devoted to the interaction of the Enterprise pilot components with the C3ISP Framework. It was possible to meet the strong confidentiality requirements of the pilot with the adoption of a centralised model, deployed in a private cloud in the trust domain of the MSSP.

Lastly, our work will continue not only in the implementation of the architecture, but also in finding significant added value services enabled by the adoption of the C3ISP Framework. As an example, malware detection techniques may benefit from the possibility of conducting collaborative analysis on datasets that are aggregated and sanitized with the C3ISP Framework. In the conclusions, this aspect is discussed in mode details.

Table of contents

1. Introduction

1.1. Overview

This document presents the first version of the design and architecture for the Enterprise pilot of the C3ISP project. The results presented here were originated from interactions within the Enterprise pilot members, but also with colleagues in the Architecture Work Package (WP7).

It is worth reporting that such interactions, especially between BT and SAP, lead to the presentation of a demo in June 2017 in an important public event (BT Innovation Week[1]) where the concept of cyber security analytics using anonymization techniques on confidential datasets was showcased and raised significant interest.

The content of the deliverable represents the basis on which the development and integration activities will take place during the next months, while implementing the first version of the Enterprise pilot by Month 24.

1.2. Deliverable Structure

The document is structured as follows:

- Section 2 presents an overview of the overall concept developed in the Enterprise pilot.
- Section 3 describes the system architecture in its different facets.
- Section 4 deepens the technical overview by proposing specific designs for each of the Use Cases presented in Deliverable D4.1.
- Section 5 analyses the data in use in the Enterprise pilot.
- Section 6 deals with the security model of the pilot, focusing in particular on authentication and authorization.
- Section 7 presents the main technical requirements for the deployment of our pilot.
- Section 8 depicts the relation between the architecture diagrams in this deliverable and the User Stories and Use Cases presented in Deliverable D4.1.
- Section 9 illustrates the integration between C3ISP Framework subsystems, existing (legacy) software in MSS Provider infrastructure and new components, developed *ad-hoc* for the Enterprise pilot.
- Lastly, Section 10 concludes the deliverable and presents our future work.

1.3. Definitions and Abbreviations

Term	Meaning
AES	Advanced Encryption Standard
C3ISP	Collaborative and Confidential Information Sharing and Analysis for Cyber Protection

[1] http://www.globalservices.bt.com/uk/en/events/innovation-week-2017

CIM	Common Information Model
CSP	Cyber Security Platform
CTI	Cyber Threat Information
DMO	Data Manipulation Operations
DPOS	Data Protected Object Storage
DPO	Data Protected Object
DSA	Data Sharing Agreement
FHE	Full Homomorphic Encryption
FMC	Fundamental Modelling Concepts
GDPR	General Data Protection Regulation (EU 2016/679), http://eur-lex.europa.eu/eli/reg/2016/679/oj
IAI	Information Analytics Infrastructure
IDE	Integrated Development Environment
IDS	Intrusion Detection System
IP	Internet Protocol
ISI	Information Sharing Infrastructure
MSS	Managed Security Services
MSSP	Managed Security Services Provider
Prosumer	An entity which is both a producer and a consumer of information, in particular of Cyber Threat Information
REST	Representational state transfer, a type of web services
SaaS	Software as a Service
SOC	Security Operation Centre
SOE	Security Operations Executive
STIX	Structured Threat Information eXpression
UML	Unified Modelling Language

2. System Overview

The main actor in the Enterprise pilot is a provider of Managed Security Operations Centre (SOC) Services to large public and private sector enterprises. This provider will be indicated in the document as MSSP as Managed Security Services Provider; the computing system offered by the MSSP is indicated as Cybersecurity Platform and abbreviated as CSP.

In the pre-C3ISP scenario, the MSSP hosts separate instances of its CSP for each of its customers. This is because of the customers' concerns about leakage of sensitive information as a result of other customers get access, directly or indirectly, to their data. Generally, the customers trust the MSSP, but are aware that mistakes and security breaches do happen.

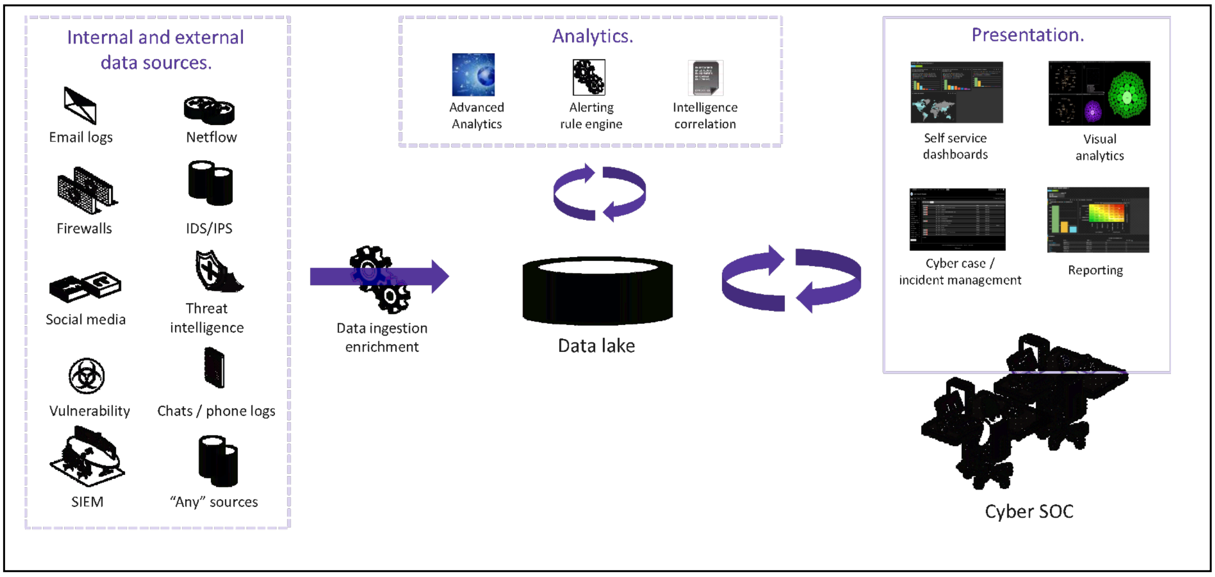

Figure 1: A single CSP instance in the pre-C3ISP scenario.

Figure 1 shows a schematic view of the structure and operation of a pre-C3ISP CSP instance. Data from a large variety of security data sources (on the left of the figure) is ingested into the platform, normalised to comply with a common information model, enriched with contextualising information and stored in a data lake. There it is available for processing by a variety of automated and man-in-the-loop analytics processes, and the results made available to human decision-makers, who are either customer personnel or MSSP personnel assigned to represent the interests of the customer in question. Analytics and Presentation tools may write results back to the data lake, so that they become available for further processing by the same and other tools. The distinction between Analytics and Presentation tool is not hard and fast — for example, visual analytics tools combine both analytics and presentation aspects.

In the C3ISP-enabled scenario, the multiple customer-specific CSP instances are replaced by a single instance serving all customers. Multiple sets of data sources are ingested by the single data lake. This is acceptable to the customers originating/owning the data because it is protected from deliberate and accidental unauthorised access by encryption and policy specification and enforcement mechanisms. Multiple teams of analysts serving the different customers now use the same platform. As well as accessing data belonging to the customer it represents, each team has policy-constrained access to sanitised data belonging to other customers, and as a result can generate improved threat intelligence. The MSSP benefits from reduced costs as a result of operating fewer platform instances, and the additional threat intelligence that can be generated by aggregating and analysing the data of multiple customers can be used to improve service and differentiate its offering from those of its competitors.

3. System Architecture

The Enterprise pilot is a centralised deployment of the C3ISP Framework (as defined in Deliverable D7.2) — single instances of the ISI, IAI and DSA Manager are installed in a data centre belonging to the MSSP, where they become part of the CSP. More details on the integration of the C3ISP Framework and the pilot software are described in the following Section 9.

The architecture of the pilot, depicted in Figure 2, comprises some elements of an existing CSP solution, coloured in *green*, others of the C3ISP Framework, in *light red* and lastly 3 new software components that are necessary for coupling the respective functionalities of these two sets. For them, it was used the *light blue*.

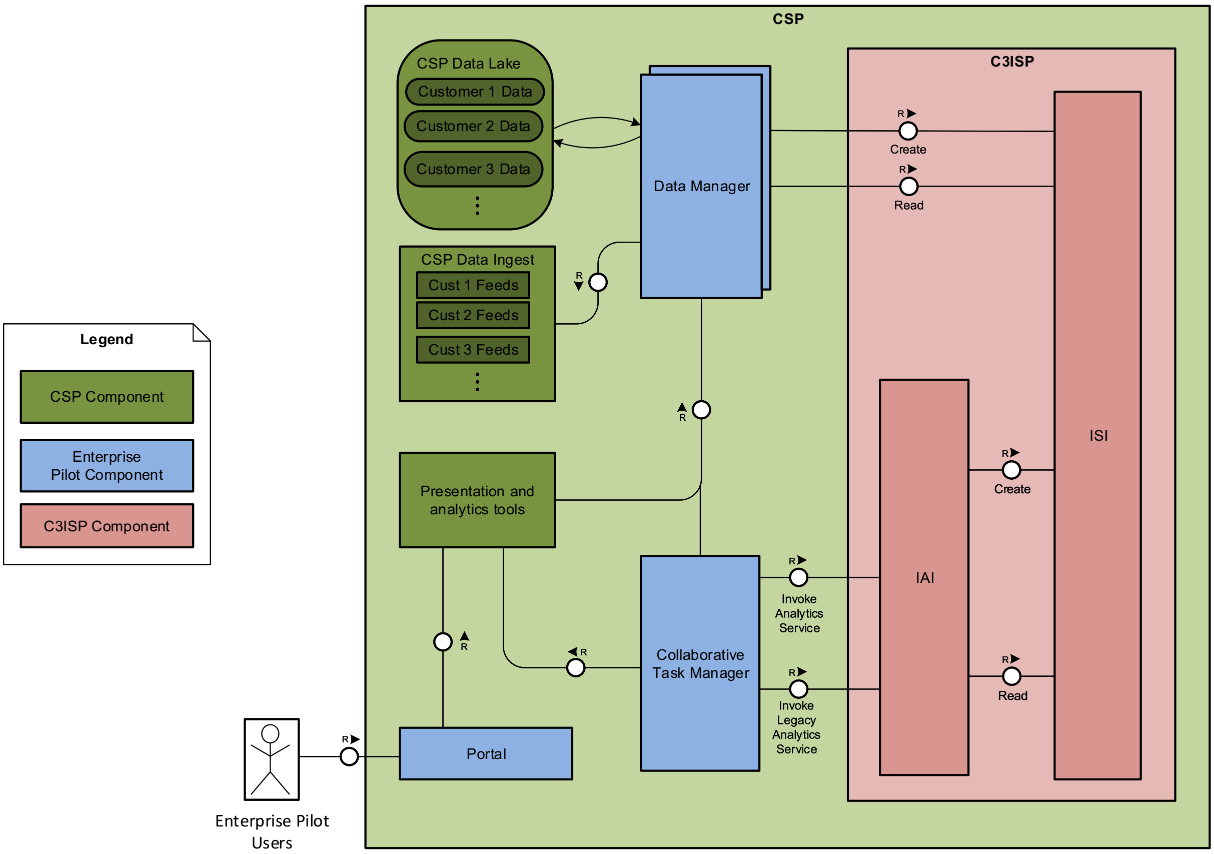

Figure 2: Software architecture of the Enterprise pilot.

Rationale for the introduction of Data Manager and Collaborative Task Manager is discussed in the following subsection 3.1 to 3.3. The Portal is a front-end that will allow all the pilot users to call their respective functionalities, just providing convenient links to the presentation and analytics tools.

3.1. Relationship between the ISI and the CSP Data Lake

The CSP Data Lake is replaced or augmented by the ISI. At this stage, the ISI functionality is not dependent on particular storage software or hardware. Consequently, the ISI's Data Protected Object Storage (DPOS, see Deliverable D7.2) could be created by integrating C3ISP subsystems with existing data lake storage resources, use newly procured storage technology specified by the MSSP, or use default technology provided by the C3ISP vendor.

The ideal case would involve complete replacement of the CSP Data Lake by the ISI, as this would mean all data would benefit from the protection offered by the C3ISP Framework.

However, situations can be envisaged where a hybrid solution would have benefits. Reasons include:

- Encryption/decryption and policy enforcement overheads could be such that only data that *needs* to undergo collaborative analytics is held in the ISI;
- A gradual migration from the pre-C3ISP to C3ISP-enabled architectures is planned;

A further possibility would be to leave the data in the CSP Data Lake, partitioned by customer (or in separate lakes), but use the ISI as a front-end, primarily to enforce data sharing agreements. The retrieved data could simply be deleted from the DPOS once accessed, or else cached for future use. This assumes that the customer enterprises are happy with the security at rest provided by the CSP Data Lake segregation. In this variant, there is little or no value in encrypting the data then decrypting again, and this may entail significant overhead. It would make sense, therefore, to turn off these operations.

3.2. The Data Manager

As introduced in Deliverable D7.2, the ISI API exposes the following operations:

- **Create** – construct a new protected bundle whose content is derived from the data provided. The data format will be normalized and data manipulation operations applied as dictated by the relevant data sharing agreement (DSA) before the result is encrypted and bundled with an appropriate policy.
- **Read** — return the contents of the referenced bundle after decryption and applying data manipulation operations applied as dictated by the policy.
- **Delete** — delete the referenced bundle if permitted by the policy.
- **Move** — cause the referenced bundle to be moved to another ISI instance. Not used in this scenario.

It is assumed that the entity invoking the operations will have to present credentials, and that the result of the operation will depend on these credentials and the relevant policy/DSA.

The ISI does not possess the full range of data management functionality required by the application, so we propose to introduce a pilot-specific Data Manager that mediates between other non-C3ISP pilot components and the ISI. For example:

- It is not possible to ask for data to be pulled in from a referenced external source.
- There is no *append* or *modify* command. To ‘modify’ an object, the old one must be deleted and a new one created. To ‘append’, a new additional object can be created, but there is apparently no way to link it to the original.

The Data Manager will implement such functionality, including:

- Maintaining a registry of external data sources corresponding to feeds from customers and third party services.
- Pulling in fresh data from these sources or from the CSP Data Lake as required and invoking the Create operation to add it to the ISI.
- Deleting bundles that are ‘stale’ or no longer required.
- Maintaining a mapping of bundles to higher-level, domain-relevant entities. An example of this is an ordered collection of bundle references mapping to a historic stream of data from a particular feed.

The Data Manager would need to be able to invoke the ISI calls with the delegated authority of various stakeholders. This will be discussed in more detail in the Security Model section below (Section 6).

3.3. Applying Analytics and Presentation tools and operations: the role of the Collaborative Task Manager

The most straightforward way to apply the CSP Analytics and Presentation tools to data stored in the ISI is for these tools to call the ISI API via the Data Manager. It is envisaged that the Data Manager would provide an API enabling it to emulate the tools' standard data source types, and that each tool instance would operate with the delegated authority of a single stakeholder.

However, there are some circumstances when invocation via the IAI offers advantages; for example:

- When authority greater or broader than possessed by any individual stakeholder is required to perform the operation on the specified data. A typical example is an operation to collect data matching a particular pattern from sources owned by different customers and return a count of the records found. No stakeholder may have the authority to retrieve all the data from the various sources, yet the result can be made widely available, as it reveals no sensitive details and it is not apparent which of the customers experienced the events.
- When there is a need to maintain a trustworthy audit trail of operations performed.

Broadly speaking one would expect pure presentation tools to interact with the ISI via the Data Manager, while operations that gather events and then summarise the aggregated information would be invoked via the IAI. However, there are many tools that fall in between these extremes, including visual analytics tools and complex automated analytics processes, so it makes sense to allow some flexibility. This flexibility could be constrained by DSAs/policies, with e.g. a data owner only allowing its data to be analysed if the relevant tool is invoked via the IAI.

It is also easy to envisage composite/ chained operations combining both modes in sequence, e.g. a 'collect and summarise' operation run by the IAI in order to generate data that is displayed to an analyst via a visual analytics tool accessing the results via the ISI.

In the same way that the Data Manager was introduced above to couple the ISI to the larger CSP, we now introduce a pilot-specific Collaborative Task Manager to integrate the C3ISP components into the CSP's analytics workflows.

4. Component Architecture

This section aims to dive down into greater level of details of the high-level subsystems identified in the previous section. In particular, the block diagrams for each pilot use case are provided using the Fundamental Modelling Concepts (FMC) notation [1] as well as the Unified Modelling Language (UML) component diagram [2] of the architecture.

4.1. Block Design

Here we expand upon the use cases (EN-UC-1, EN-UC-2 and EN-UC-3) described in detail in the Deliverable D4.1 [3].

4.1.1. EN-BD-1: Block Design from EN-UC-1

This use case concerns a security analyst, an employee of MSSP, whose role is to identify, analyse and investigate actual and potential threats to the security of a number of assigned customers of the MSSP. He/she is responsible to keep the CSP Knowledge Base up-to-date with the latest types and methods of attacks such that they can be detected automatically in the future. Figure 3 shows the FMC block diagram of this use case.

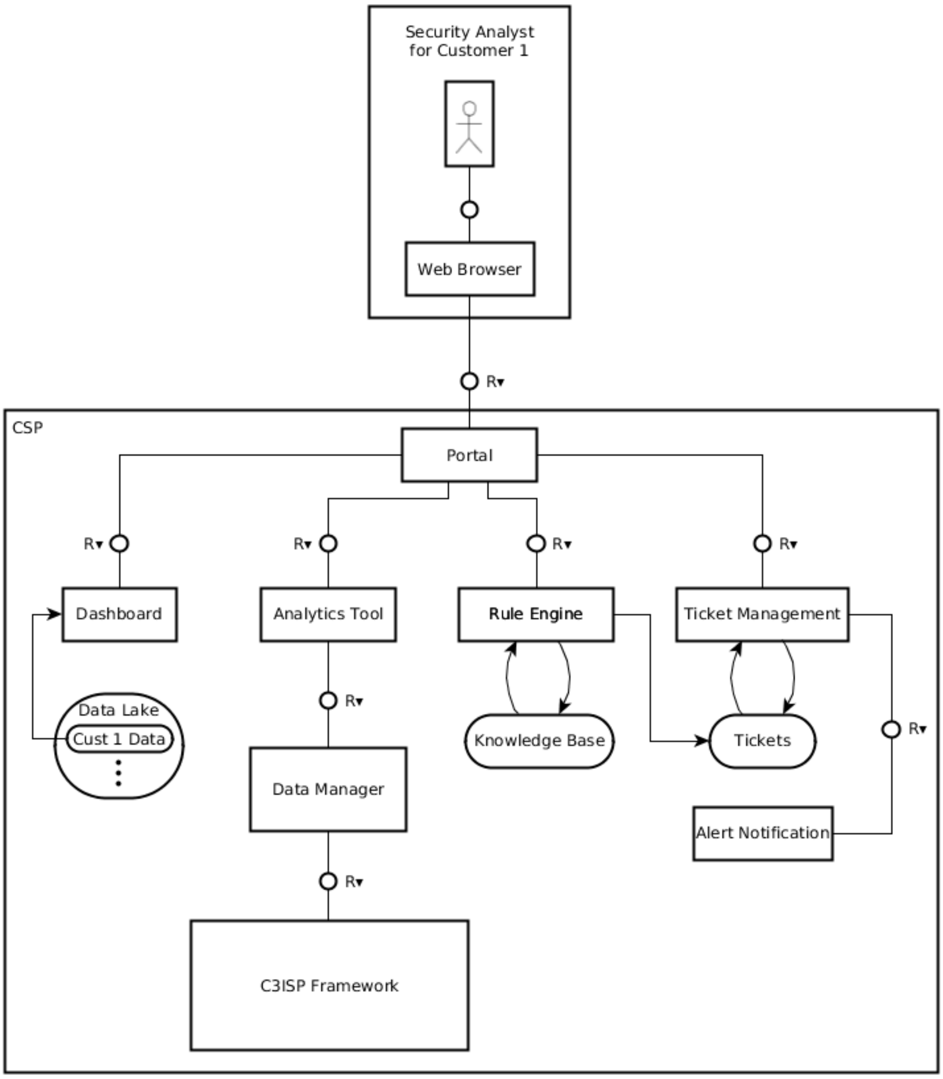

Figure 3: FMC block diagram for the 'Identify new threat' use case (EN-UC-1).

The Security Analyst uses a Web Browser to log in to the Portal which provides access to different components of CSP, i.e., Dashboard, Analytics Tool, Rule Engine, and (Trouble) Ticket Management. The Dashboard reads the data from the CSP Data Lake and provides a customised view of various security metrics (e.g., top 10 blocked source IPs), as well as, current threats and anomalies in the environment of each customer (e.g. Customer 1). The Analytics Tool is used for in-depth investigation of potential new threats and anomalous events. It queries the relevant data from the C3ISP Framework via the corresponding Data Manager instance. The C3ISP Framework will ensure that the read operation complies with the corresponding DSA policy and the data is formatted correctly before it is passed to the Data Manager. To help the investigation, the Security Analyst may request any related data or threat intelligence (i.e. CTI) shared by other customers via the C3ISP Framework. The DSA policy of each requested CTI will be enforced including any data sanitisation process (e.g. anonymization) to satisfy the confidentiality and privacy requirements. After completing the investigation, the Security Analyst may decide to update the CSP Knowledge Base by specifying new rules or patterns for the Rule Engine. The Security Analyst can then log in to the Ticket Management system to create or open a new ticket to alert the customer about the threat via the Alert Notification system, e.g. by e-mail. In the future, such tickets will be generated automatically by the Rule Engine.

4.1.2. EN-BD-2: Block Design from EN-UC-2

This use case concerns a Data Policy Officer of an enterprise customer who wants to specify sanitization measures, access and usage control directives and other means proposed by the C3ISP Framework in order to lower the sensitivity of the data of her/his employer prior to its sharing in the CSP. Such policies are stored as Data Sharing Agreement (DSA). The FMC block diagram of the use case in shown in Figure 4.

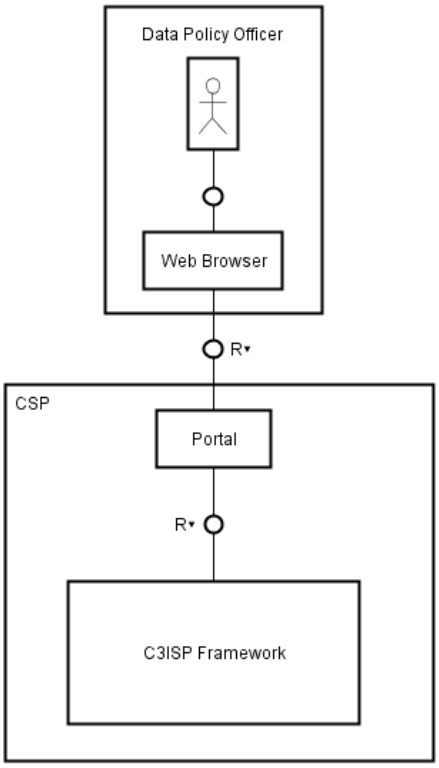

Figure 4: FMC block diagram for the "Define data sharing policy" use case (EN-UC-2).

The link to the DSA Editor tool (as part of the C3ISP Framework) may be provided by the Portal of the CSP. The Data Policy Officer uses a Web Browser to log in to the Portal and use the DSA Editor for writing the DSA policies for his/her enterprise data. The DSA policies are then persistently stored in the C3ISP Framework.

4.1.3. EN-BD-3: Block Design from EN-UC-3

This use case concerns a Security Operations Executive (SOE) of an enterprise customer who wants to use the analytics capability of the CSP on their own enterprise data as well as on aggregated data from multiple enterprise customers if available. Figure 5 shows the FMC block diagram of the use case.

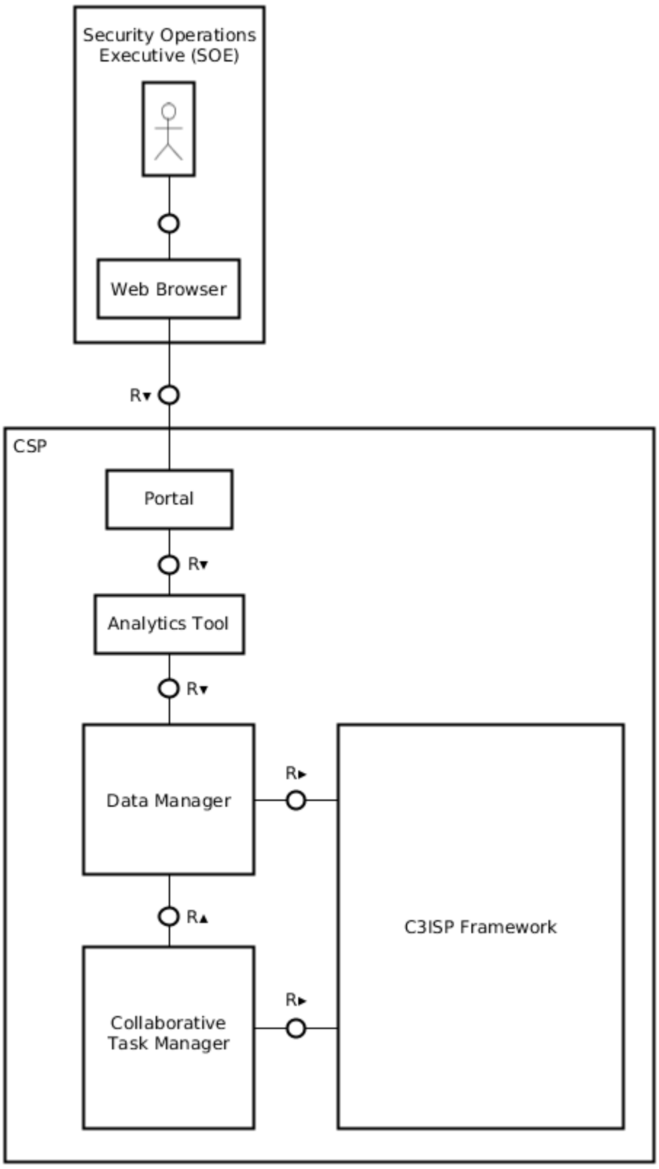

Figure 5: FMC block diagram for the "Analyse enterprise security data" use case (EN-UC-3).

The SOE uses a Web Browser to log in to the Portal, which provides a link to the Analytics Tool component. The Analytics Tool retrieves the SOE's enterprise data (i.e. CTI) via the corresponding Data Manager instance, which in turn sends the read request to the C3ISP Framework. After the data object is fetched from the C3ISP Data Protected Object Storage (DPOS) and the associated DSA policies checked, the data is reformatted and passed to the Data Manager which forwards it to the Analytics Tool. The Collaborative Task Manager is communicating with the C3ISP Framework to request its collaborative analytics service (provided by the C3ISP Analytics Engine). The Collaborative Task Manager may request specific information from the Data Manager about the customer data that needs to be aggregated with other customers' data for the purpose of collaborative analytics, e.g., to verify the occurrence of a suspicious source IP in other customer's network traffic. It then sends a request for such analytics function to the C3ISP Framework, which first retrieves the relevant data (e.g., list of CTIs) from its DPOS, performs the function on aggregated data from multiple enterprise customers and stores the analytics result as a new CTI in DPOS. This new CTI can then be requested by the Data Manager instance and consumed by the Analytics Tool. It is anticipated that the Collaborative Task Manager will carry out its task with elevated privilege.

4.2. Component Design

This section presents a high-level component diagram for the pilot. An accurate component diagram would require a deep dive into a number of legacy components that are currently deployed as part of the actual CSP. However, due to the need of easing the adoption of C3ISP contributions into existing systems, a number of different integration strategies are currently under study to maximize the stated objective. For example, the Data Manager interactions with legacy analytics may use different interfaces, from approaches *a la* SQL to plain file exchanges. Deliberately, it was decided not to make any binding decision at this stage but to leave this decision at the implementation phase, after an evaluation of the different options. Reports on this matter will appear in the upcoming deliverables.

4.2.1. EN-CD-1: Main component design diagram

For the mentioned reason, the component diagram in **Figure 6** remains at a level of details similar to the block diagrams of the previous Section 4.1, with similar explanation. Accordingly, it is possible to notice the role of the Data Manager and Collaborative Task Manager as mediators of the interactions between the legacy components (only represented in a subset) and the C3ISP Framework.

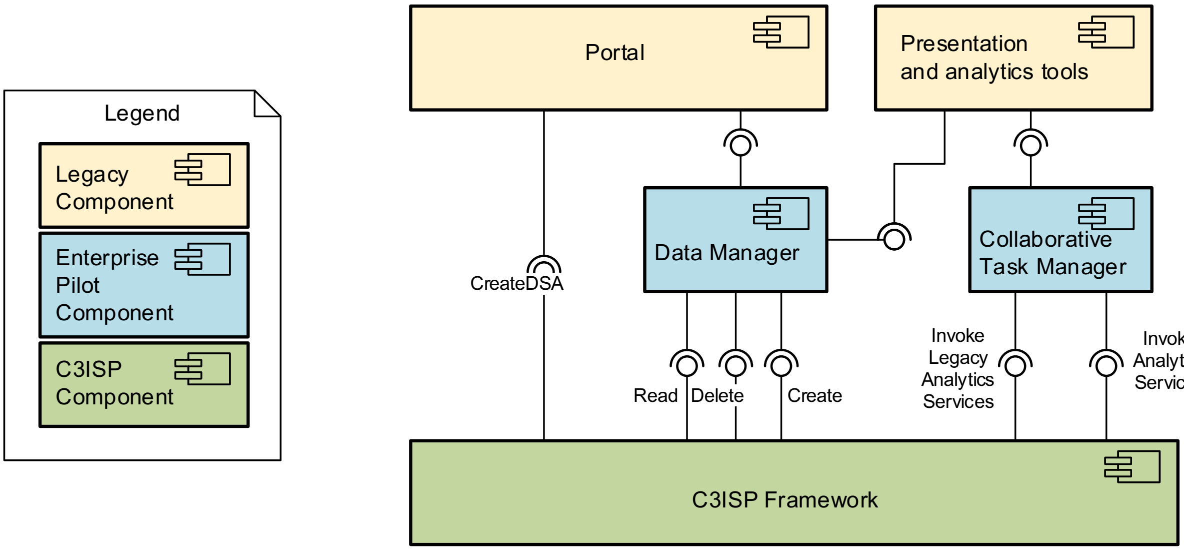
Legend
Legacy Component
Enterprise Pilot Component
C3ISP Component
Portal
Presentation and analytics tools
Data Manager
Collaborative Task Manager
CreateDSA
Read
Delete
Create
Invoke Legacy Analytics Services
Invoke Analytics Services
C3ISP Framework

Figure 6: Component model for the Enterprise pilot.

5. Data Model

Various events and logs are collected from the enterprise customer environments and ingested by the CSP, which then analyses them for patterns of activity that indicate malicious or undesirable behaviour from a security perspective. The CSP consists of several analytics engines that are designed to perform different types of analysis on various inputs including Intrusion Detection System (IDS) alerts, firewall logs, and host system logs. As pictured in Figure 2 all customer data are stored in CSP Data Lake. The Data Manager would then pull fresh data from CSP Data Lake – either regularly or on-demand, depending on the data sharing policy – and invoke the *Create* operation to add it to the ISI.

The ISI is designed to process input data coming in different formats and convert it into a structured CTI (e.g. in STIX[2] format). Nevertheless, the MSSP usually has already a framework in place to normalise the data into a common model of key-value pairs before enriching it with contextualising information and ingesting it into the platform. This enables the CSP to take structured and unstructured event data from various security products that customers may have deployed in their environment and pass it on to the analytics and presentation tools without requiring vendor-specific configurations, e.g., the same set of rules for detecting *Denial of Service* attacks can be applied to the data of different customers. It is anticipated that some security vendors may have adopted a standard format for their security events such as the Common Event Format (CEF) [4], which would facilitate the normalisation process according to the MSSP's own Common Information Model (CIM). Figure 7 shows the whole data ingestion process, starting from its collection until it is added to the C3ISP's ISI subsystem.

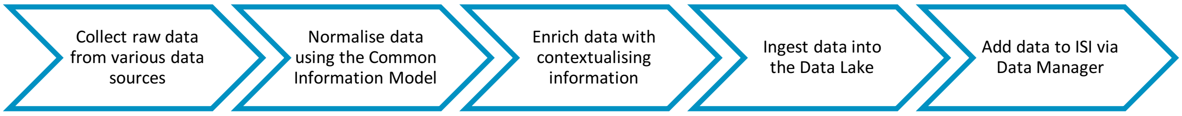
Collect raw data from various data sources
Normalise data using the Common Information Model
Enrich data with contextualising information
Ingest data into the Data Lake
Add data to ISI via Data Manager

Figure 7: The data ingestion process.

The CIM provides a range of data models that allow security and analytic use cases to be realised. Each data model specifies a set of keys (i.e., field names) to represent an actual event of a domain of interest. Typical data models include *Intrusion Detection*, *Malware*, and *Network Traffic*. The CIM is an internal specification used by the MSSP, and its details cannot be disclosed in this document. To implement the Enterprise pilot, we plan to derive data models from a publicly-available CIM, such as, the Splunk CIM[3]. Table 1 shows an example *Intrusion Detection* data model that describes the events gathered from IDS devices. Data models for other types of events, e.g., malware attacks or network traffic, may also be derived and used later during the implementation depending on the availability of the relevant datasets.

Table 1: Intrusion Detection data model

Field name	Data type	Description	Possible values
action	String	Action taken by the IDS	
category	String	Vendor-provided category of the	

[2] https://oasis-open.github.io/cti-documentation/stix/intro

[3] http://docs.splunk.com/Documentation/CIM/4.8.0/User/Overview

		attack event	
date_time	String	The date and time of the attack event	
dest	String	Destination IP of the attack	
device_name	String	Hostname of the device that detected the event	
device_ip	String	IP address of the device that detected the event	
ids_type	String	Type of IDS that generated the event	network, host, application, wireless
severity	String	Severity of the event	critical, high, medium, low, informational
signature	String	Vendor-provided signature name of the attack event	
src	String	Source IP identified in the attack event	
user	String	The user involved in the attack event	
vendor_product	String	The vendor and product name of the IDS or IPS system that detected the attack event	

6. Security Model

The security model of the Enterprise pilot foresees the adoption of authentication and authorization measures to protect customer data. This section illustrates such measures.

To this extent, it is useful to consider the actors that are part of the Enterprise pilot. It is possible to distinguish them in two different groups:

- MSSP employees
    - Analysts
    - Account Manager
    - Developer Manager
- Customer's employees
    - Security Operations Executive
    - Data Policy Officer

6.1. Authentication

The authentication needs of human actors of the Enterprise pilot will be addressed with the same mechanisms. However, specific authorizations will be assigned to each of them. Authentication/authorization will be necessary:

- To allow human actors to interact with the Portal.
- To allow the machine-to-machine interactions among Portal, Data Manager or Collaborative Task Manager, and C3ISP Framework.

Table 2: The security model of the Enterprise pilot

	Authentication	Communication Encryption	Authorization
Human to Portal (or Analytics tool)	HTTP Basic Auth (see Section 6.1.1)	HTTPS (SSL)	ABAC (see Section 6.2)
Machine to Machine	OAuth + IdP (see Section 6.1.2)	TLS/SSL secure protocol	see Deliverable D7.2 Section 7.1

6.1.1. Authentication from Human to MSS Portal

The security analysts of the Enterprise pilot may access their functionalities through the Portal, using a browser. The data exchange is secured in transit through HTTPS protocol.

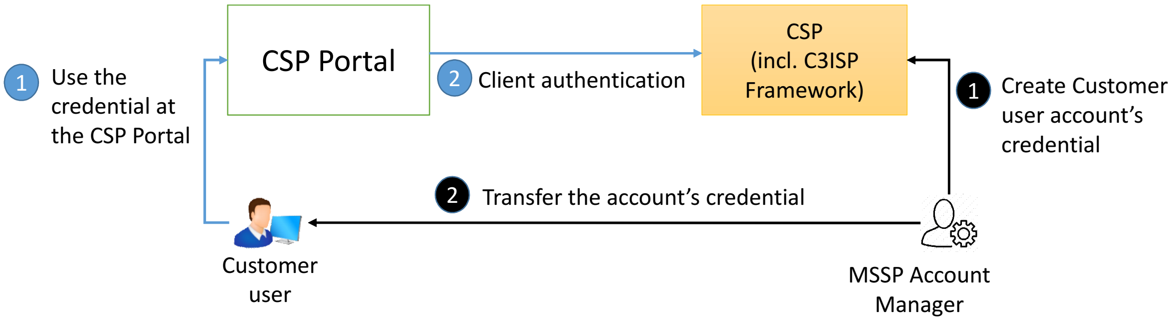
1 Use the credential at the CSP Portal
CSP Portal
2 Client authentication
CSP (incl. C3ISP Framework)
1 Create Customer user account's credential
2 Transfer the account's credential
Customer user
MSSP Account Manager

Figure 8: HTTP Basic Authentication from Customer user to CSP: the black lines refer to configuration and the blue lines refer to authentication.

To do so, their accounts must be created in the CSP. The user provisioning procedure changes according to the affiliation of the considered user. If it is a new customer user (as depicted in Figure 8):

1) MSSP Account Manager creates a user account credential (username and password). Other attributes can be specified, e.g. the user role.
2) MSSP Account Manager transfers the credentials to the Customer user via a secure channel.

In case the user is an employee of the MSSP, the procedure is similar but it is enacted by the Developer Manager.

The HTTP Basic Auth that allows CSP backend services (integrated with C3ISP Framework) to authenticate a Portal user, with the following steps:

1) The user credentials are provisioned in the Portal web page rendered by the browser.
2) The browser interacts with the CSP that authenticates the active user by means of the account credentials.

Note that after the authentication, the user's commands can be transferred from the Portal to the CSP.

6.1.2. Authentication Machine to Machine

In order to authenticate the different components that are part of the Enterprise pilot, a flexible approach will be adopted, to allow the integration of existing CSP components with the C3ISP Framework. This choice is facilitated by the C3ISP Framework design decision to apply an Identity Manager component for all the communications happens via web services that need to be protected for authentication and authorization (see Deliverable D7.2 Section 7.1). As a well-known standard that can be used also for API protection, C3ISP refers to OAuth2, which is an authorization framework that integrates with different authentication mechanisms. More details about this can refer to Deliverable D7.2, Section 2.1 and 7.1.

6.2. Authorization

6.2.1. Authorization of CSP and C3ISP Framework

CSP uses a legacy access control mechanism to protect its functionalities. In this section, we describe the relevant authorizations for the Use Cases specifically addressed in the Enterprise pilot. Such authorizations will be enforced by means of the C3ISP Framework capabilities. An Attribute-Based Access Control (ABAC) approach will be pursued, for facilitating the required authorization granularity. This approach permits to consider as input for authorization decision, an arbitrary combination of attributes of the requestor, the environment and the target resource.

Once the client side is authenticated (normally in the Portal), the authorizations are verified for when the user requests an operation, for example through the Data Manager. Some authorizations come from the DSA associated to the resource(s), others are assigned to the user role. The formers are specified by each customer's Data Policy Officer through the DSA Editor tool, while the latter are as follows.

- MSSP employees

    - The Analysts are allowed to access a customer's data through the analysis tools when they are assigned to work on that customer, unless otherwise specified by the Data Policy Officer.
    - The Account Manager deals with customer user provisioning and can see CTI and reports made available by an Analyst in the context of an incident.
    - The MSSP Developer Manager has administrative access to the system.
- Customer's employees
    - The Security Operations Executive follows up to incidents reported by the Analyst, having access to CTI and reports made available through the CSP Portal by an Analyst.
    - The Data Policy Officer defines the directives to be applied to his/her employer's data, by means of the DSA Editor tool.

7. Requirements for Testbed Environment

All backend components of the Enterprise pilot are deployed locally on an integrated cloud-based testbed environment of the MSSP. The testbed consists of non-C3ISP Framework components such as the CSP Data Lake, Data Manager, Collaborative Task Manager, the Analytics and Presentation tools, as well as the C3ISP Framework subsystems, such as, the DSA Manager and the ISI and IAI subsystems. As mentioned in Section 3 the Data Manager and the Collaborative Task Manager are introduced in the Enterprise pilot to mediate between the C3ISP Framework and other CSP legacy components such as Data Lake, Analytics and Presentation tools. The hardware and software requirements for the C3ISP Framework subsystems will be similar with the ones of the C3ISP Framework (see Deliverable D7.2 for details). In the following, the minimum requirements for non-C3ISP Framework components are listed.

7.1. Hardware Requirements

7.1.1. Hardware Requirements for CSP Data Lake

The hardware requirements to run the Data Lake are as follows:

- Processors: 1x Intel/AMD 64-bit (Quad-core)
- Minimum RAM: 8GB
- Hard Disk: 200GB

7.1.2. Hardware Requirements for Analytics and Presentation Tools

A visual analytics tool and a rule engine are planned to be deployed on the testbed. Other components such as authentication server, Portal (and Dashboard, one of its components) will also be deployed to some extent. Their hardware requirements are given below:

- Processors: 2x Intel/AMD 64-bit (Quad-core)
- Minimum RAM: 16GB
- Hard Disk: 200GB

7.1.3. Hardware Requirements for Data Manager and Collaborative Task Manager

The hardware requirements to deploy the Data Manager and the Collaborative Task Manager are as follows:

- Processors: 1x Intel/AMD 64-bit (Quad-core)
- Minimum RAM: 8GB
- Hard Disk: 100GB

7.2. Software Requirements

7.2.1. Software Requirements for CSP Data Lake

The software required to install and run the Data Lake are as follows:

- Operating system: Linux, e.g. Ubuntu distribution

- Apache Hadoop system, e.g. Cloudera's open-source Hadoop distribution (CDH)[4]

7.2.2. Software Requirements for Analytics and Presentation Tools

The visual analytics tool (SATURN) and the Rule Engine are both stand-alone web applications that have been developed by BT Research & Innovation. The visual analytics tool will be deployed "as is" while further development is planned for the Rule Engine and Portal to suit the pilot's use cases. We anticipate making use of the following software/tools on the testbed (including requirements from SATURN and the Rule Engine):

- Operating system: Linux
- Database: PostgreSQL and Oracle
- Other software: Apache Tomcat, CAS (Central Authentication Service)[5], Elasticsearch[6], Kibana[7]

7.2.3. Software Requirements for Data Manager and Collaborative Task Manager

The Data Manager and Collaborative Task Manager will be developed as web service applications. The software framework and technology used for their development will later be decided during the implementation phase. Nevertheless, we anticipate making use of the following software/tools:

- Operating systems: Linux
- Database: PostgreSQL or Oracle
- Other software: Apache Tomcat, Apache Axis2

[4] https://www.cloudera.com/products/open-source/apache-hadoop/key-cdh-components.html

[5] https://www.apereo.org/projects/cas

[6] http://elastic.co/products/elasticsearch

[7] http://elastic.co/products/kibana

8. Requirements Matrix

The following matrix summarises how the Use Cases (and the associated User Stories) presented in Deliverable D4.1 [3] can be implemented adopting the design presented in the previous sections.

Table 3: Mapping Use Cases and User Stories to architecture diagrams

Use Cases	User Stories	Description	Coverage
EN-UC-1	EN-US-1	The analyst must be able to detect new threats by analysing all data made available by all customers that opted-in for such possibility.	The diagram **EN-BD-1** permits to combine CSP data collection and legacy analytics software with the C3ISP Framework ISI and IAI thus offering to the analyst the union of all necessary information of the different customers who allowed for this possibility in their DSAs.
EN-UC-2	EN-US-2	The Data Policy Officers of customers must be able to define their own DSAs.	The diagram **EN-BD-2** shows how Data Policy Officers, through the Portal, may get access to the DSA Editor thus permitting the definition of DSAs for their data.
EN-UC-3	EN-US-3	Customers' Security Operation Executives must be able to assess the cyber-security risk of their companies, also taking advantage of analysis made using sanitized data of other customers that opted-in for such possibility.	The diagram **EN-BD-3** fulfils the requirement by means of Data Manager and of Collaborative Task Manager: in this way, with the help of the former, sanitized data are retrieved and with the latter, arbitrary analytics may be performed, still in compliance with DSAs specified by data owners.

9. Integration Plan with C3ISP Architecture

In order to cope with the necessary confidentiality requirements stemming from the relation binding the MSSP with its customers, the Enterprise pilot need to consider a model where the ISI and IAI of the C3ISP Framework are deployed in the trust domain of the MSSP, as illustrated in Deliverable 4.1.

Moreover, as information collected in all customers' networks are already collected in data lakes under the control of the MSSP, it seems sufficient to dispose of single instances for ISI and IAI. The deployment of DSA Editor in MSSP trust domain is deemed not critical and any option is viable, provided that information confidentiality in transit is preserved.

Taking into account these aspects, the integration plan selected for the Enterprise pilot is the 'fully centralised' option, with a notable difference with respect to the definition provided in Deliverable D7.2.

In the diagram of Figure 9, elaborated from Deliverable 7.2, the colour *green* indicates the trust domain where the C3ISP Framework is visible. The C3ISP Framework is considered to be deployed on-premises or anyway in a private cloud managed by the MSSP.

Considering the Enterprise pilot operations, and in particular EN-UC-1 and EN-UC-3, this would allow for a simpler management of the 'virtual data lakes' in which sanitised data of different customers will be made available, on request.

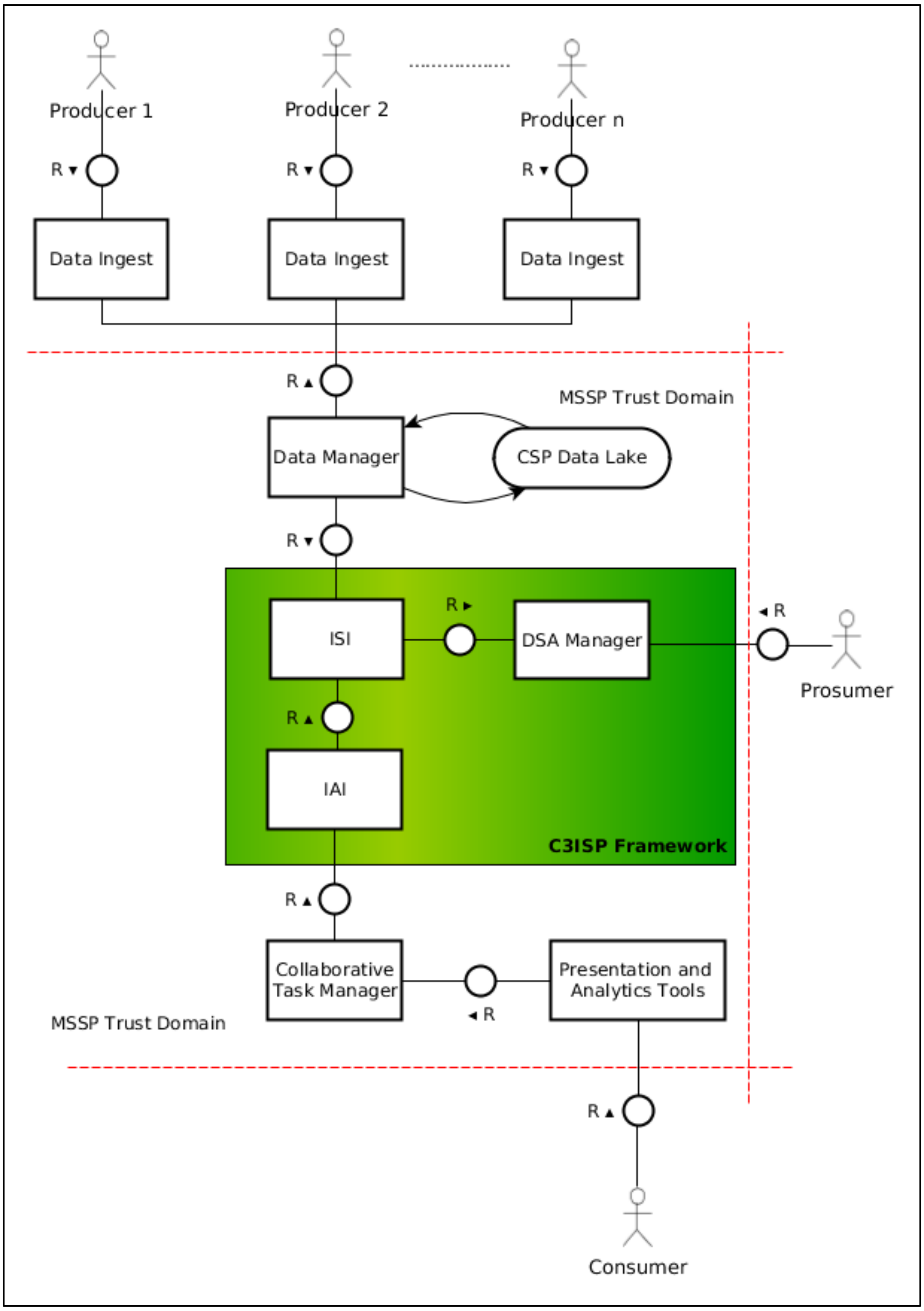

Figure 9: The 'fully centralised' deployment of C3ISP Framework architecture, as defined in Deliverable D7.2 but with the C3ISP Framework deployed in a private cloud.

10. Conclusions and Future Work

This document presented the initial architecture for the C3ISP Enterprise pilot. The complexity of this scenario is represented by the need to integrate different C3ISP contributions in a complex environment, where analysts develop new practices on a daily basis in order to counter the always mutating cyber threats.

The significant amount of data processed daily by the MSSP is another challenge, still minor with respect to the impact C3ISP contributions will need to make in the eyes of CSP customers in order to allow the usage of their data for cross-customer analysis.

Still, the concepts and architecture developed so far seem to integrate well with the existing MSSP infrastructure. Specific components must be introduced (Data Manager and Collaborative Task Manager) to act as "adapters" between legacy and C3ISP subsystems, in order to reconcile the expectations of legacy components with the new analysis capabilities of the IAI or the possibility to access sanitised data lakes for all customers' data.

In the upcoming months, we will focus on fine-tuning the integration among legacy components and C3ISP Framework, using the approaches described in this document. However, we will not neglect to pursue concrete use cases, like those presented at the BT Innovation Week 2017, for example considering advanced malware detection techniques.

New malwares, such as Wannacry, NotPetya and Emotet, pose a huge threat to enterprise security. Industry has seen these new malwares propagate from one business sector to another in an unprecedented speed with uncountable cost and damage [5][6]. This is because these new malwares are equipped with both internal and external network propagation capabilities [7]. The C3ISP Framework, as integrated in the Enterprise pilot, would help to address this malware propagation issue from threat discovery to early warning and control:

- C3ISP would allow data from participating enterprises to be shared for malware intelligence discovery. The unknown/unclear malware threat pathway from one business sector to another could be uncovered using the shared information.
- C3ISP would make performing malware analytics at a large scale possible and easier. Malware data mining and machine learning require large amount of data. C3ISP would help develop in-depth malware analytics using advanced data mining and machine learning techniques on a large amount of data and subsequently improve each participating enterprise's ability to protect their own network.
- C3ISP would enable detection of malware threat in one participating enterprise to be shared with another enterprise. Early warning of malware threat could be triggered and automatically sent to the participating enterprises.

11. References

This section lists the references used throughout the document:

[1] Apfelbacher, R., Rozinat, R.: Home of Fundamental Modelling Concepts – Notation Reference, http://www.fmc-modeling.org/notation_reference, last accessed: 2017/09/28.

[2] UML, Unified Modeling Language, Component Diagrams, http://www.uml-diagrams.org/component-diagrams.html, last accessed: 2017/09/28.

[3] Di Cerbo, F. (ed.): Requirements for the Enterprise Pilot, C3ISP Deliverable D4.1, 2017.

[4] Common Event Format (CEF), White Paper, ArcSight, Inc., 2010.

[5] https://www.theregister.co.uk/2017/08/16/notpetya_ransomware_attack_cost_us_300m_says_shipping_giant_maersk/, last accessed: 2017/09/28.

[6] https://www.theregister.co.uk/2017/09/20/fedex_notpetya_damages/, last accessed: 2017/09/28.

[7] https://www.fidelissecurity.com/threatgeek/2017/07/emotet-takes-wing-spreader, last accessed: 2017/09/28.