Advanced Cyber-Threat Intelligence, Detection, and Mitigation Platform for a Trusted Internet of Things
Grant Agreement: 786698

D7.5 CYBER-TRUST information and evidence storage

Work Package 7: Distributed ledger technology for enhanced accountability

Document Dissemination Level

P	Public	☒
CO	Confidential, only for members of the Consortium (including the Commission Services)	☐

Document Due Date: 31/01/2020
Document Submission Date: 31/01/2020

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

Document Information

Deliverable number:	D7.5
Deliverable title:	CYBER-TRUST information and evidence storage
Deliverable version:	1.0
Work Package number:	WP7
Work Package title:	Distributed ledger technology for enhanced accountability
Due Date of delivery:	31/01/2020
Actual date of delivery:	31/01/2020
Dissemination level:	Public
Editor(s):	Clément Pavué (SCHAIN)
Contributor(s):	Nicholas Kolokotronis (UOP), Kostantinos Limniotis (UOP), Sotirios Brotsis (UOP), Olga Gkotsopoulou (VUB), Dimitris Kavallieros (KEMEA)
Reviewer(s):	Olga Gkotsopoulou (VUB) Dimitris Kavallieros (KEMEA)
Project name:	Advanced Cyber-Threat Intelligence, Detection, and Mitigation Platform for a Trusted Internet of Things
Project Acronym	Cyber-Trust
Project starting date:	1/5/2018
Project duration:	36 months
Rights:	Cyber-Trust Consortium

Version History

Version	Date	Beneficiary	Description
0.1	17/01	SCHAIN	Initial version
0.2	19/01	VUB	Enhanced legal section
0.3	26/01	UOP	Enhanced security section
0.4	28/01	SCHAIN	Final version sent to reviewers
0.5	30/01	KEMEA-VUB	Feedback
1.0	30/01	SCHAIN	Final version for submission

Acronyms

ACRONYM	EXPLANATION
API	Application Programming Interface
C&C	Command and Control
CA	Certificate Authority
CI / CD	Continuous Integration / Continuous Delivery
DDoS	Distributed Denial of Service
DLT	Distributed Ledger Technology
HTTP	Hyper Text Transfer Protocol
IDS	Intrusion Detection System
IoT	Internet of Things
ISP	Internet Service Provider
LEA	Law Enforcement Agency.
OS	Operating System
SDA	Smart Device Agent
SGA	Smart Gateway Agent
SOHO	Small Office / Home Office
TTL	Time To Live
UI	User Interface
URL	Uniform Resource Locator
TxGen	Transaction generator
EvGen	Evidence Generator

Executive summary

One of the responsibilities of the Cyber-Trust Distributed Ledger Technology (DLT) is to allow partners to safely share information about alleged malicious activity between them. In this context safe means, that information is guaranteed to circulate without any modification from malicious sources. This will speed up the investigation of Law Enforcement Agencies (LEAs) as they will be able to access information remotely and not only physically.

Table of Contents

Table of Figures

1. Introduction

1.1 Purpose of the document

The main objective of the deliverable is to provide a technical documentation of the second prototype of the DLT implementing the forensic evidence and its sharing between organization member of the Cyber-Trust program.

One of the focus of the deliverable is the measures taken by the consortium of Cyber-Trust to comply with the legal framework when creating, storing and sharing forensic evidence.

The second focus is the methods used by Scorechain to implement the measures previously described in the document.

1.2 Relations to other activities in the project

As the DLT is used by most the components developed for the Cyber-Trust project, the outcome of this deliverable will most likely impact them. The most impacted components are the Network and Asset repository [A16] as this component is responsible to push metadata of the new evidence to the DLT. Another component impacted is the Visualization Portal [A01] as it will render the forensic evidence stored inside the DLT.

1.3 Structure of the document

The document has two axes. The first one is the decision taken by the consortium to fit inside the legal framework when creating, storing and sharing metadata of material that may contain forensic evidence. The second axis is the technical details of the implementation of the measures previously taken consortium-wise.

2. Sharing forensic evidence metadata through the DLT

The legal implications with regard to the storage of personal data on DLTs have been studied in detail in D3.2 and in particular, Chapter 6. The report covering the issue of admissibility of electronic evidence, addressed, among others, two main questions: 1) what are the requirements for the lawful collection of electronic evidence and its respective admissibility before a court? and 2) is a DLT-based chain of custody possible? As for the first question, given the European aspiration of the project, the document concluded that any tools for the collection of evidentiary material should follow the legal requirements at EU level and the principles described both in (European Union Agency on Cybersecurity) ENISA´s and Council of Europe´s guides on electronic evidence as well as the national framework of the jurisdiction where the evidence is collected from and the jurisdiction where it is going to be used for the criminal proceedings. Concerning the second question, if a DLT was to be used in the chain of custody, then a private solution seemed appropriate to be prioritized, supported by the necessary security safeguards.

Cyber-Trust's goal is to address the challenges, derived from a smart home (or SOHO) environment, concerning the collection and preservation of forensic evidences by taking advantage of the technology of IDSs and distributed ledgers. A number of tools installed at the getaway of a smart home, authorize the monitoring of interconnected devices, the identification of known threats and zero-day vulnerabilities as well as the collection of forensic evidence for detected malicious activities and traffic. Each malicious interaction is securely stored in an off-chain database maintained by the ISP while the metadata and the hashes are stored on the DLT. Thus, Cyber-Trust platform makes it possible for LEAs not only to trace back the sequence of malicious interactions to their source but also to realize the *chain-of-custody*.

2.1 Security components

A smart home domain is comprised by the following security components, which are illustrated in **Error! Reference source not found.**:

- *Smart Gateway Agent (SGA)*: The SGA is an essential component that is responsible for the security of the smart home network. This component utilizes innovative intrusion detection approaches to activate the monitoring activities regarding the health of the environment and characterizes the behaviour of the IoT devices as well as the collection of forensic evidences. Upon the registration of a device, in which the device is in a clean state, the *SGA* extracts the behavioural patterns of the device by performing device fingerprinting and identifying the suspicious traffic.

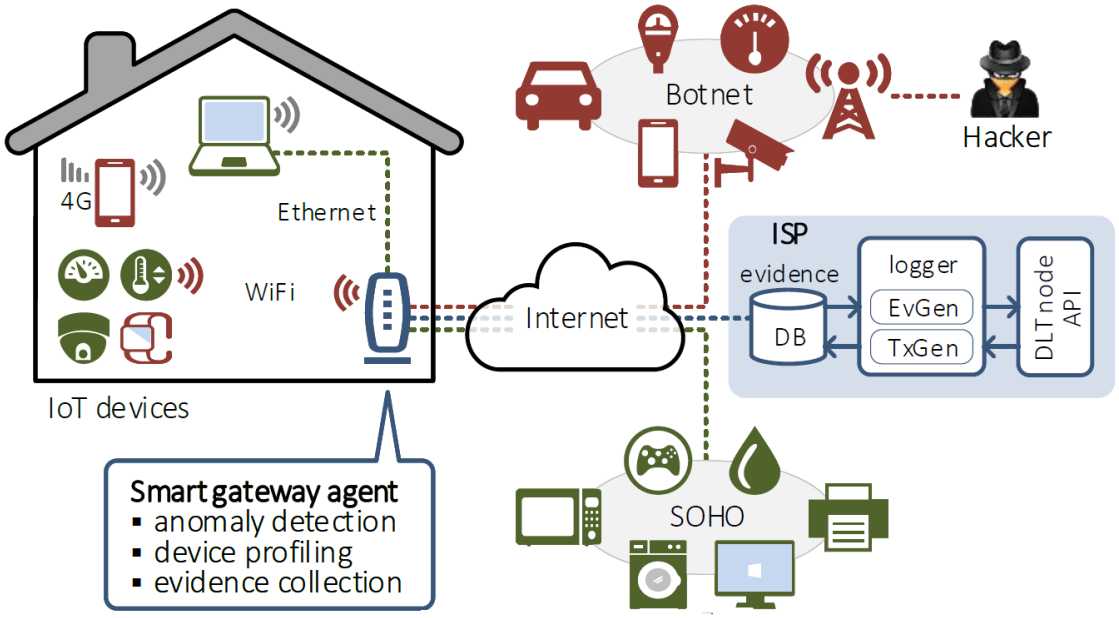

Figure 2-1: Cyber-Trust's view for the forensic evidence collection process

- *Smart Device Agent (SDA)*: In more capable devices (e.g. smartphones, devices with more computational power) an *SDA* is installed to enable the acquisition of evidences from the IoT devices. The SDA runs in a restraining manner while its responsibilities are to monitor not only the security of the device (vulnerabilities, patching statues, integrity) but also the device's usage and all the critical files.

2.2 Collection and erasure of forensic evidence

Upon the detection of suspicious network traffic or any malicious device activity from the SGA or the SDA (respectively), the ISP receives the collected evidences and stores them as raw data in the off-chain database (based on the national legislation and the applicable safeguards). The process is intended to confirm the integrity and secrecy of the forensic evidences upon transferring and storage. Furthermore, the Cyber-Trust platform ensures that only secure systems are authorized to collect and receive the evidences. To achieve that, a trust relationship by means of an attestation protocol is established to identify the configuration (such as the MBR, firmware, BIOS) of the remote device. As illustrated in Figure 2-1, a proof-of-existence of the forensic evidence is attained using the DLT.

Upon the creation of new evidence log by the *logger*, the collected evidence is inserted into the off-chain database (*evidence DB*) – this is indicated with *EvGen* in Figure 2-1. For this step to be accomplished, (i.e. to digitally sign the new evidence), the *logger* has generated a key pair using a digital signature algorithm, which in our case is the certificate authority (CA) of HyperLedger Fabric. Thus, upon the generation of digitally signed evidence, a new identifier is also created in order to confirm the distinctiveness of the new evidence's identifier. After the creation of the log event, each proof of existence of the forensic evidence is written on the DLT by means of a sequence of transactions – this is indicated with *TxGen* in Figure 2-1.

It is eminent that to erase a digital evidence the authorized entity has to create a new transaction, but the metadata concerning the specific evidence cannot essentially be eliminated from the DLT, as this would indicate that the entire mechanism would have to be altered. It is only possible for authorized entities to erase an evidence from the off-chain database and then by creating a new transaction, state that the evidence does not exist anymore. This solution was the closest to erasure, that could be identified by the technical partners at this stage, in order to correspond to the relevant data protection legal requirements (in particular, Article 17 of the General Data Protection Regulation and Article 16 of the Law Enforcement Directive). The extent to which, this solution is satisfactory should be considered in relation to the current state of the art and the continuous legal debate concerning the legal interpretation of 'erasure'.[1] Further, in the light of the Directive which covers the processing of personal data for criminal proceedings, the data subjects' rights with respect to personal data contained in a judicial decision or record or case file (Article 18 of the Law Enforcement Directive), may be carried out in accordance with national law. This entails that the implementation and enforcement of data subjects' rights, including the rights to erasure and rectification as well as the respective data controllers' obligations, may differ from state to state.

2.3 Approach taken by Cyber-Trust on DLT

Provided the legal limitations and uncertainties with regards to the new technologies used in Cyber-Trust, and under the vision of a proof of concept, the technical partners have chosen to proceed with the following solutions:

- A private and permissive ledger is used.
- On-chain will only go the metadata of the material which is allegedly relating to a cyber-attack.

[1] For an extensive analysis on the topic, please see the recent publication of the European Parliament Research Services (EPRS) on "Blockchain and the General Data Protection Regulation - Can distributed ledgers be squared with European data protection law?", pp.74-78. https://www.europarl.europa.eu/RegData/etudes/STUD/2019/634445/EPRS_STU(2019)634445_EN.pdf

- Access to the on-chain metadata will be provided only after a successful LEA's request, which would satisfy all the legal requirements per jurisdiction.
- The actual forensic evidence will remain exclusively on the ISP side, off-chain and will be accessed only after all the legal requirements have been satisfied.

When the partners will need to share information inside the Cyber-Trust platform as it has been established, the consortium members have assessed that the only way to ensure the integrity and the reception of such information is through the DLT. Some of the information is not personal data and as such, can be added on-chain. This is the case of the list of patches published by an organization. However, material that may contain or point to electronic evidence is not meant to be shared with every partner, thus a series of measures have to be taken to enhance compliance with privacy and data protection obligations.

2.3.1 Confidentiality level

The first measure was to define the confidentiality level of each class defined in the data model of the DLT introduced in D7.2. Three levels of confidentiality were created; public information available to every partner of the Cyber-Trust platform. This information is the one introduced in the previous deliverable of Work Package 7 relative to trusted file storage. Next category is private information that can't be shared between partners. It corresponds to the data responsible of the ownership management. This information determines if the request for modification on a device by another person than the owner of the device can be executed based on the authorization level of the requester. The last section is private information that can be shared between partners. It concerns metadata referring to forensic evidence and other information useful to LEAs during criminal proceedings. At their creation, data is private and only visible to the creator of the data. Please find below an updated version (see D7.3 for previous version) of the data model with regards to the confidentiality model.

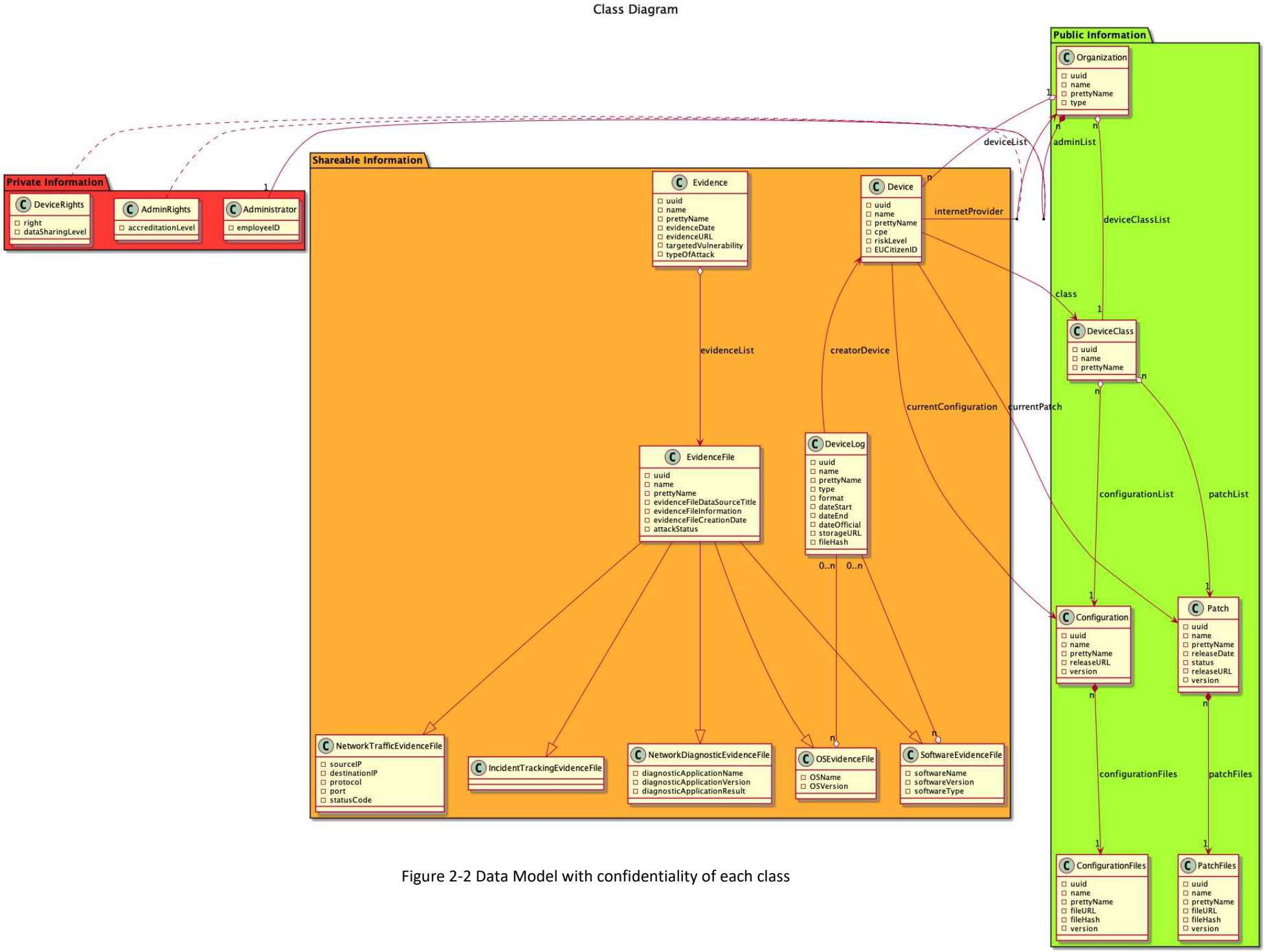

Figure 2-2 Data Model with confidentiality of each class

2.3.2 Communication in the DLT

The next measure was to establish how the partners will communicate with each other to share the information. To do so, three solutions were proposed.

1. DLT is not used in the communications between organisations.

Communication between partners remains as is. This was not a real option to us. It was proposed in case a legal issue arise or communicating with the help of the DLT was not suitable to them.

2. Communications information is stored in the DLT.

Each organization upon registration provides an email address to be contacted if needed (for instance legal@isp.com). Email addresses are just stored to provide integrity, ensuring you're contacting the right person. In this option, they are used as support to initiate communication. After reaching an agreement on the information to share, the owner of the data uses the API of the DLT to share the data. However, this solution doesn't keep tracks of the content of the conversation.

3. Communication occurs on the DLT.

Communications between two partners are stored in the DLT. This information will be stored as private information that can't be shared on both of the peers owned by the communicants. The resulting data will be useful as Cyber-Trust will have a complete backlog of the communications. In that case, DLT is used as a way of communication. Content of the message will be text and PDF files.

The solution chosen based on the end-user's feedback (LEAs) was the third one. The contacted LEA agreed that the DLT should be used in order to provide integrity and ownership of requests and at the same time, to preserve the transparency and history of communications, for establishing a chain of custody. They also recognize that the DLT will be used in the chain of actions for the LEAs.

2.3.3 The Time To Live (TTL) Feature

The LEAs were concerned, however, on the duration the information will be available on their backend as they can't store for ever data related to forensic evidence. So, the last measure taken was to add a way to delete automatically data after a specified time stored on the DLT. This concerns the data that can be shared between partners. As the time to live may differ in different legal frameworks, the TTL is passed in the DLT through the API with other parameters at the creation of the data. In the mock-up we used TTL of five years for the data shared between the peers, as an indicative timeframe, but the time can be adjusted to the particular needs of each jurisdiction. This mean that after five years of storage inside the DLT, the data will be automatically deleted.

3. Technical implementation

3.1 Changes since D7.3

As stated in Section 2 of the document, the DLT will be used to store messages and files that are circulated between the LEAs and the ISPs when the first are asking to access forensic evidence metadata and consequently, the actual evidence file.

For the sake of the mock-up the number of peers of the DLT were changed. Cyber-Trust had three peers. One for the Cyber-Trust backend responsible for the propagation of the transaction. The remaining peers were two organizations-partners of Cyber-Trust. Cyber-Trust now has five peers. One is still dedicated to the central backend. Two are for peers of ISPs and two for LEAs. As collections of data were created with different confidentiality, it was necessary to differentiate whether the peer is hosted by an ISP that can create data related to forensic evidence or an LEA that can request access for forensic evidence. The APIs were splatted so as a peer can only request what it has access to. This has been done as part of the devOps effort.

Regarding the tools and technologies used for the development of the DLT, same as in D7.3 have been used. However, since a new major release of HyperLedger Fabric was disclosed on the 12th of December 2019, it was necessary to study the advantages and the drawbacks of updating the version of HyperLedger Fabric from 1.4 to 2.0-beta. The major advantage is the time, the Docker images built from the new version of Fabric is lighter due to the use of Linux Alpine instead a security oriented, lightweight Linux distribution. The deployment of our containers would have been faster, so it would ease the development of our solution. The update was carried out as part of the process of the continuous improvement of the platform.

3.2 Implementation of the data model

As part of the last development iteration the data model was created and the smart contract was written, along with the respective unit tests so as to ensure their robustness. The missing part of the data model was the deviceLogs, messages as well as the forensicEvidence part. Our API with its Swagger documentation was updated with regards to the update of the implementation of the data model.

The major changes since D7.3 is the use of private data, due to the sensitivity of some information stored by the DLT. This requires additional work to enable the use of private data inside a DLT using HyperLedger. Indeed, if a node only uses public data, the communication happens on channel. A channel is a “subnet”[2] of the network. Usually one public channel is used by all the peers to circulate the public information and multiple channels are created by peers to circulate information to authorized peers. To do so, the storage of the node is performed by two entities, the **Public Block Storage** and the **Public State Database(s)**, as drawn on the figure below.

[2] https://hyperledger-fabric.readthedocs.io/en/release-1.4/channels.html

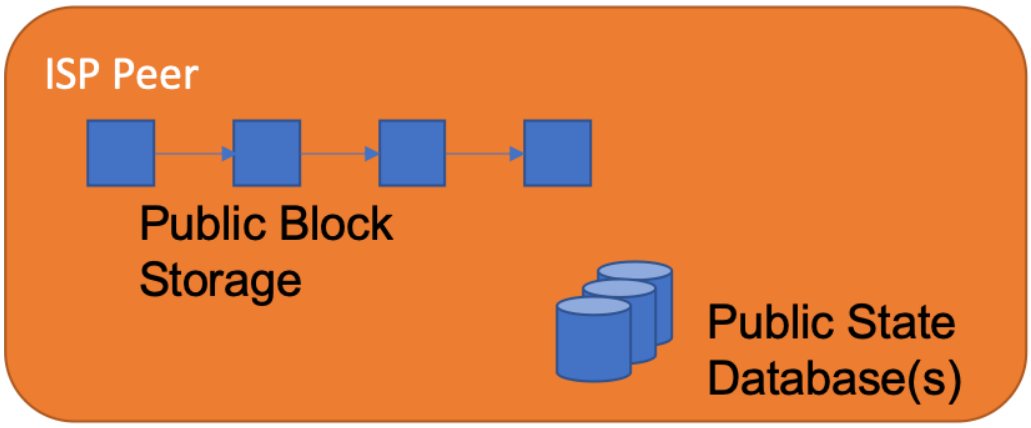

Figure 3-1 Architecture of an ISP peer without private data

Note that the Figure 3 represents the peer run by an ISP but the same applies for a peer run by a LEA. The **Public Block Storage** is where the peer keeps all the public transactions made on the Blockchain. There is only one instance of **Public Block Storage** no matter the number of channels the node is subscribed to. The **Public State Database(s)** stores the state of each channel separately, so if the peer is subscribed to N channels, it will run N instances of **Public State Database**. Creating a new channel requires a lot of DevOps and is hard to maintain if many channels are created.

This is why it was decided to take advantage of the recent addition of the private data feature. The first change with private data is that a single channel can handle multiple private collections. For instance, if we take the network of our mock-up, the communication between peers is the following:

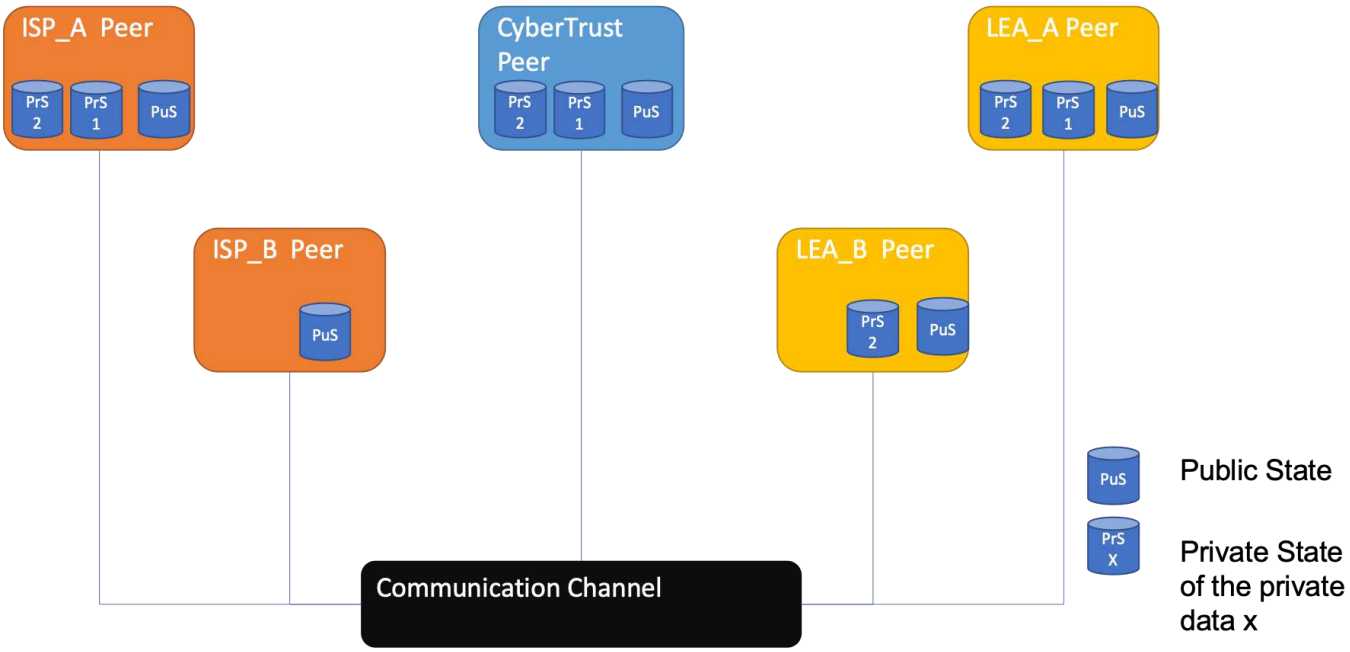

Figure 3-2 Architecture of private data communication

As it is depicted in Figure 3-2 two peers belong to ISPs, two to LEAs and the last one is hosted by the Cyber-Trust organization. Only one channel is used for all the communications between peers. All nodes have access to the data stored publicly by Cyber-Trust partners. These data are located in the public storage, PuS (see Figure 4). On the same figure, two private data collections, PrS1 and PrS2, have been created. Each collection

refers to the creation of forensic Evidences related to an attack the ISP has encountered. So here, ISP_A has created two private data collections, related to two different attacks. PrS2 has been shared to both LEAs' peers that requested through the communication mechanism described in Section 2 of this document. PrS1 is only shared to LEA_A, as LEA_B didn't request it. Now let's see how a peer stores private data.

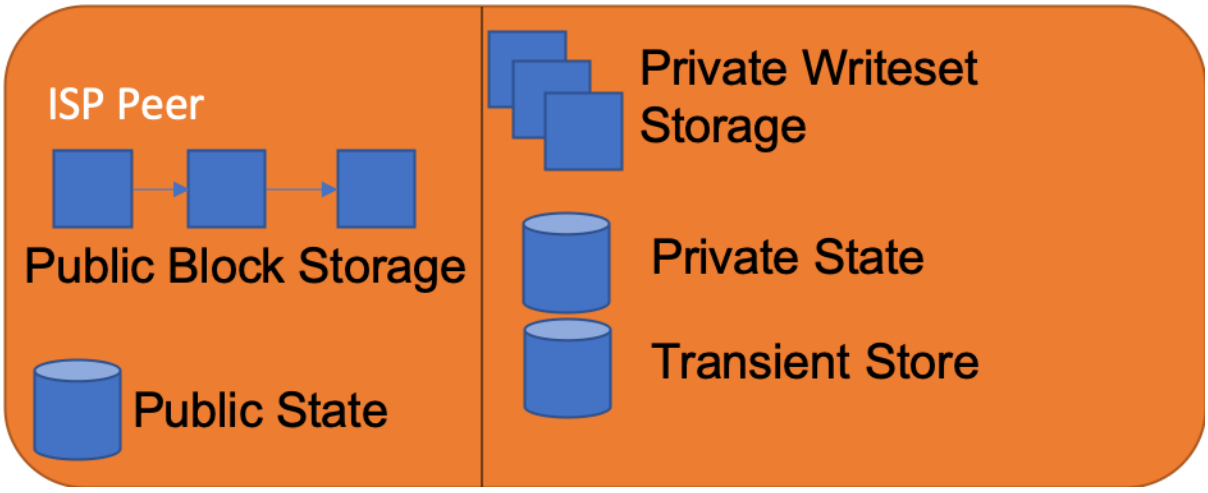

Figure 3-3 Architecture of an ISP peer with private data

The node keeps the same element as in Figure 3 to store public data. For private data, three more elements are necessary. **Private Writeset Storage** stores the content of the private data. One instance is needed for each Private data collection the peer has subscribed to. This element behaves as a traditional database. **Private state** behaves like **Public State** but for the data stored privately. Figure 3-4[3] depicts what a peer stores, depending on whether it is authorized or not to access the data.

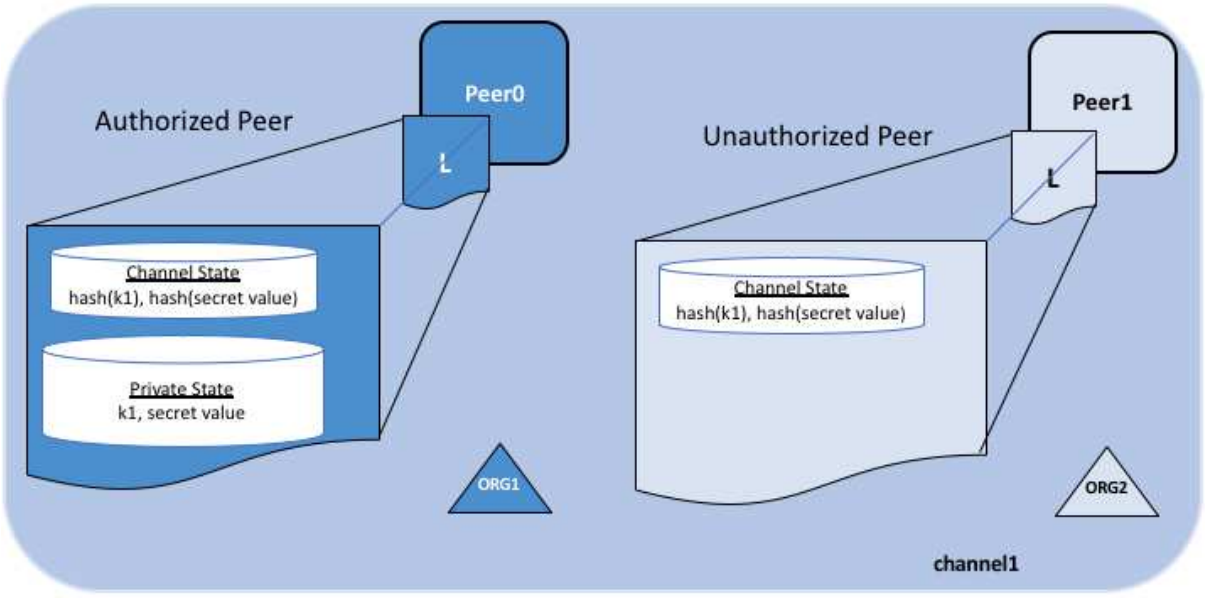

Figure 3-4 Content of a peer with private data

Channel State refers to **Public State Database** of Figure 3. The unauthorized peer can see the existence of a transaction, validate its state with the hash but can't access its content.

[3] https://hyperledger-fabric.readthedocs.io/en/release-1.4/private-data/private-data.html

Based on Figure 6, you can see that an unauthorized peer can see that a transaction related to private data was added to the ledger. To enhance the privacy of the data submitted and to comply to the legal framework, Cyber-Trust DLT's API will only display the transaction with private data a peer has access to.

3.3 DevOps

To improve our development experience and ease our partners to deploy and use the DLT, a part of our work for this iteration includes DevOps tasks that will be depicted in this section. The first one, was to configure the automatic deployment of our peer on OTE testbed. As it was presented in D7.3, Docker and Kubernetes were used to deploy the DLT on OTE testbed. The use of these technologies was already saving the developer time when testing the DLT on OTE testbed. To be production-ready though, it was needed to automatically deploy the DLT, for the partners to receive automatically a new version of it, with potential bug fixes, additional features and dependencies updates. It has been configured in two branches of Gitlab, master and develop. The develop branch is the environment which host the code that hasn't been tested yet. To test the latest advancement of the DLT and execute the unit test, especially the test of the smart contract. This is a guarantee that there is no regression introduced on the code. On master branch, is located the code that is used for the mock-up, demonstration to members of the Cyber-Trust consortium, this is also the branch where partners will pull the code. So, the same testing procedure is applied. Any partners that pull the open-source code of the DLT on the master branch will receive automatically updates of code with the help of a Gitlab runner.

With the creation of the data confidentiality model, different roles for the peers have been created. Indeed, a peer hosted by a LEA and a peer hosted by an ISP don't have the same features enabled. This differentiation is made at the time of deployment. Thus, partner can't change their role by modifying the code themselves. For the sake of our mock-up, each peer it is accessed by a different URL that takes this form www.{organization-name}.{branch-name}.cyber-trust.neofacto.eu (www.isp1.master.cyber-trust.neofacto.eu) . Each organization is represented by a Kubernetes pod which encloses all the needed containers for it to work independently: a UI with the enabled features, a Backend, the blockchain peer, a couchDB data base that stores the blockchain state. Additionally, with the use of KubeDB it is ensured that every one of these organizations has its own Postgresql data base. All these components communicate via configured services and are exposed to URLs through ingresses.

3.4 Visualization of the forensic evidence

In an effort to keep the interface of Cyber-Trust tightly coupled, it has been decided at the consortium level that Scorechain will implement in its API, the necessary endpoints to visualize the forensic evidence. The API will return the data as JSON. Thus, Scorechain will provide MATHEMA, the partner responsible of the Visualization Portal[A01], with the code to convert the JSON to HTML.

4. Conclusion

This document presented the measures taken to enhance the communications between Cyber-Trust partners and the share of data. These measures have been implemented during the second development iteration of the DLT, in which all the end user requirements regarding the DLT (as described in WP2) have been met. These measures are mainly: the use of the DLT as a way for partners to communicate and the storage of information as private data.

The next step for the DLT will be to improve its security framework with regards to the scalability.