E-CORRIDOR
Edge Enabled Privacy & Security Platform For Multi Modal Transport

D7.1

Data Analytics Techniques Requirements and Architecture

WP7 – Data Analytics techniques

E-CORRIDOR

Edge enabled Privacy and Security Platform for Multi Modal Transport

Due date of deliverable: 31/05/2021
Actual submission date: 31/05/2021

31/05/2021

Version 1.0

Responsible partner: UTRC
Editor: Stefano Sebastio
E-mail address: stefano.sebastio@rtx.com

Project co-funded by the European Union 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 E-Corridor Project is supported by funding under the Horizon 2020 Framework Program of the European Union DS-2018-2019-2020, GA #883135

Authors: Stefano Sebastio, Riccardo Orizio, Amine Lamine (UTRC), Thanh-Hai Nguyen, Hoang-Gia Nguyen (CEA), Ilaria Matteucci, Gianpiero Costantino, Giacomo Giorgi, Andrea Saracino (CNR), Ruisong Han (WIT), Roland Rieke, Florian Fenzl (FhG), Roghayeh Mojarad, Koussaila Moulouel, Abdelghani Chibani (PEC)

Approved by: Roland Rieke, Florian Fenzl, Christian Plappert, **(**FhG), Patrizia Ciampoli (HPE)

Revision History

Version	Date	Name	Partner	Sections Affected / Comments
0.01	16-Feb-2021	S. Sebastio, T.-H. Nguyen	UTRC, CEA	Initial table of content
0.02	30-Mar-2021	I. Matteucci, G. Costantino	CNR	Contribution to T7.1, T7.3 and T7.5 on driver identification and intrusion prevention system
0.03	02-Apr-2021	R. Orizio	UTRC	Contribution to T7.1 on passenger location
0.04	04-Apr-2021	R. Han	WIT	Contribution to T7.2 and T7.4 on CO2 aware itinerary planning
0.05	08-Apr-2021	R. Rieke, F. Fenzl	FhG	Contribution to T7.5 on automotive intrusion detection
0.06	09-Apr-2021	A. Lamine	UTRC	Contribution to T7.1 passenger identification and contextual analysis
0.07	12-Apr-2021	G. Giorgi	CNR	Contribution to T7.1 gait analysis
0.08	15-Apr-2021	R. Mojarad, K. Moulouel, A. Chibani	PEC	Contribution to T7.1 face and activity recognition
0.09	27- Apr-2021	G. Giorgi, A. Saracino	CNR	Contribution to Sec 7 – pilot specific analytics ISAC
0.10	28-Apr-2021	T.-H. Nguyen	CEA	Contribution to T7.3 on OpenAPI for fully homomorphic encryption
0.11	02-May-2021	T.-H. Nguyen	CEA	Contribution to T7.5 on homomorphic encryption-based intrusion detection
0.11	04-May-2021	S. Sebastio	UTRC	Introduction, matching to use cases, contributions to project objectives and conclusion
0.12	06-May-2021	S. Sebastio	UTRC	Completed WPL review and submission to internal reviewers
0.13	18-May-2021	S. Sebastio	UTRC	Integration of the review comments from FhG and HPE
0.14	29-May-2021	R. Orizio, I. Matteucci, G. Costantino, R. Rieke, R. Han, A. Lamine, H.-G. Nguyen, T.-H. Nguyen, S. Sebastio	UTRC, CNR, FhG, WIT, CEA	Improvements according to the feedback of the internal reviewers

Executive Summary

This document contains requirements and architecture for the data analytics techniques constituting the toolbox of the Information Analytics Infrastructure (IAI) subsystem of the E-CORRIDOR framework. The data analytics components included in the toolbox and their requirements have been designed and/or tailored to fulfill requirements and needs of the use cases identified by the three pilots of the project, namely AT (Airport-Train), S2C (Smart cities and car sharing) and MMT-ISAC (Multi-Modal Transportation Information Sharing and Analysis Center).

Each component in the toolbox is discussed by considering the current state of the art (with the aim of identifying the technology gaps), its characteristic features and the planned maturation in E-CORRIDOR. Moreover, data and component requirements with respect to the ones of the architecture, identified in D5.1 (“Requirements for the E-CORRIDOR architecture") for the first milestone of the project at month six (M6), are reported. The requirements elicitation process carried out by the above mentioned three project pilots at M6 (respectively in deliverables D2.1, D3.1 and D4.1 – “Requirements for the Pilot”) are exploited here to show how the analytics proposed for the toolbox meet the pilots’ needs, are critical for the depicted scenarios and are therefore well-placed for integration in the multi-modal transportation domains. This has the twofold goal of easing further maturation of the data analytics components and (possibly) generating products for the pilots. Given the close collaboration of all the partners in project consortium, during the meetings some opportunities for synergies among analytics and advanced security services (through the ASI subsystem, WP8) have been identified and are presented here.

Table of contents

1. Data Analytics Techniques

The E-CORRIDOR framework offers to its (data) *prosumers* (i.e., producer and consumer) a set of data analytics for (cyber-) security and advanced services tailored for the multi-modal transportation entities and their users. The whole framework, as described in detail in D5.2 ("First version of E-CORRIDOR Architecture"), is composed of five main subsystems, denoted with blocks of different colors in Figure 1.

The data analytics components constitute the *analytics toolbox* (see Figure 1 – component 1 marked in red) of the Information Analytics Infrastructure (IAI) subsystem of the framework and are the focus of this deliverable. Through classification and prediction (e.g., through machine learning), additional knowledge is extracted from the data collaboratively shared in the Information Sharing Infrastructure (ISI) subsystem by the data producer.

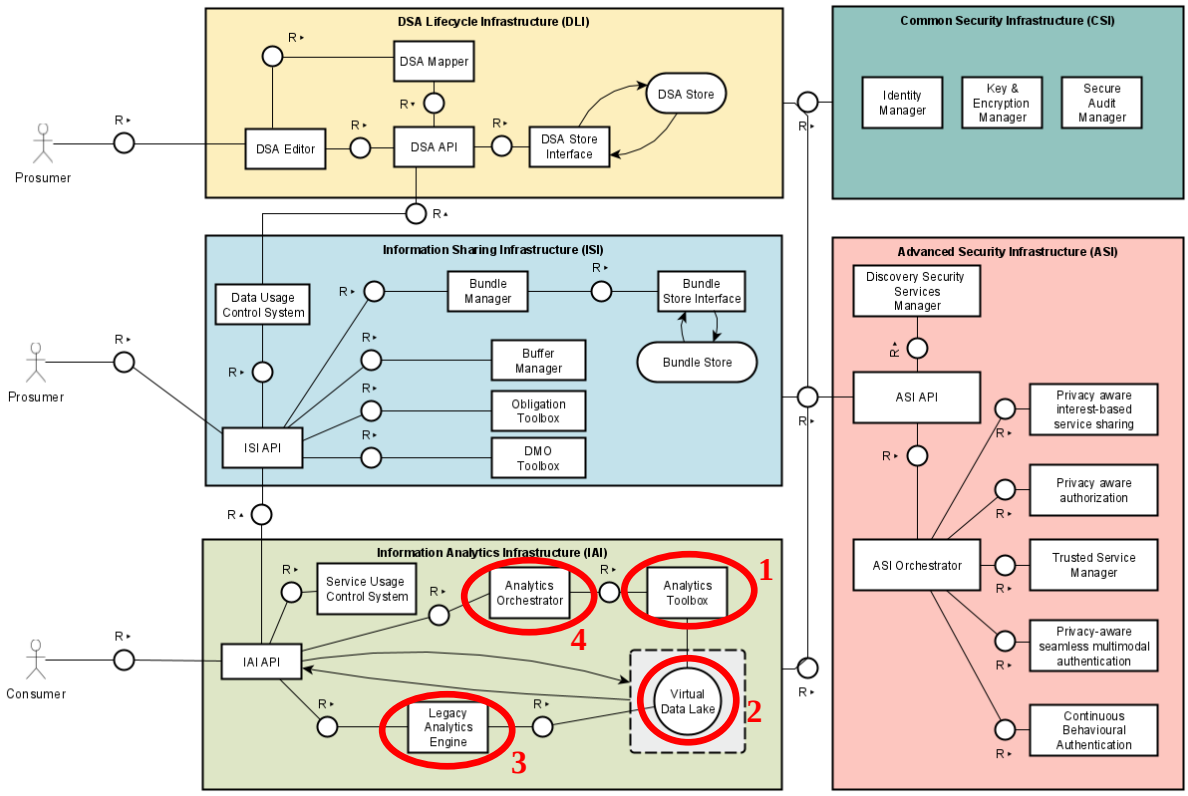

Figure 1 Architecture of the E-CORRIDOR framework (with red ovals some components mentioned in this section are highlighted). Please refer to D5.2 for a detailed description of the framework and its subsystems.

Here are briefly reported the project objectives (Obj.) to which the activities carried out in Work Package (WP) 7 – "Data Analytics Techniques" contribute: (Obj. 2) define edge-enabled data analytics and prediction services in a collaborative, distributed and confidential way; (Obj. 3) define a secure and robust platform safe from cyber-attacks and able to ensure service continuity; (Obj. 4) improve, mature and integrate existing tools provided by the partners and tailored to needs of platform and pilots; (Obj. 5) provide mechanisms for a seamless access to multimodal transport; (Obj. 6) deliver pilot products.

By taking into account the objectives of the E-CORRIDOR project (discussed in detail in Section 9) and to cover all the broad set of requirements expressed by the project pilots (Obj.

6) in the spectrum of privacy, security (Obj. 3), and services for the multi-modal transportation entities and users (Obj. 5), several analytics have been included in such a toolbox. Examples of such data analytics identified and designed, at the time of writing this deliverable (month twelve of the project, M12) are aimed at passenger and driver identification, itinerary planning, security and intrusion technologies. Over the course of the project, new analytics could be identified, and for the ones described in this deliverable it is expected a maturation (Obj. 4) in terms of accuracy, efficiency, novel features and capability to be executed in a hybrid edge-core cloud architecture (Obj. 2).

Any stakeholder in the multi-modal transportation domains defined by the project pilots, including their users (e.g., driver or passenger) and external (authenticated) source of data can contribute to the collective generation of the data made available in the ISI and later exploited by the analytics. Therefore the data that the latter have to process are heterogeneous and span from video-camera feeds, CAN bus (controller area network) messages, OBD (on-board diagnostics) readings, IMU (inertial measurement unit) sensor data, RSSI (received signal strength indicator), travel preferences and directions just to mention a few.

1.1. Structure of the Deliverable

The remaining of this deliverable is structured as follows. In the following of this section, the flexibility embedded in the data analytics toolbox is described (see Section 1.2). Then, the main interactions of such toolbox with the E-CORRIDOR framework are discussed (see Section 1.3). Finally, the adopted naming convention (see Section 1.4) for the analytics identified at the time of writing this deliverable (and listed in Section 1.5) is introduced.

Sections from 2 to 6, detail on the data analytics components available in the toolbox. The division in sections corresponds to the logical grouping presented above (and matches the tasks described in the Description of the Action, DoA, of the E-CORRIDOR project). After a brief overview providing a brief outlook on the context and summarizing the main features, for each component, the current state of the art is reviewed. Then, characteristics of the components, their expected functioning and adoption in the project pilots (including data input and output format requirements) are discussed. The description of each analytics component ends with tables summarizing their requirements (and the match to the ones identified for the E-CORRIDOR platform at M6 and reported in D5.1), the application to the pilot use cases and potential synergies identified with other analytics and advanced security services (described in D8.1 “Advanced Security Services requirements and architecture”). Pilot specific analytics (in particular for the MMT-ISAC) are reported in Section 7.

To remark the applicability of the described analytics and their importance for the pilots, in light of a successful demonstration and potential exploitation as products, Section 8 summarizes the match of the pilot use cases with the analytics components presented here.

Finally, the contribution of the analytics to the objectives of the E-CORRIDOR project are discussed in Section 9 and conclusions are in Section 10. Bibliographic references are reported in Section 11. Acronyms used in the document and details on the data formats used by some of the analytics are reported in the Appendix.

1.2. A Plug-in Approach for the Analytics in the Toolbox

The analytics toolbox is constituted by components deployed either as executable Java (JAR) [1] or as services packaged in software containers (e.g., with Docker [2]). By supporting a *plugin* approach, new analytics components can be *dynamically* added to the toolbox without any change required on the E-CORRIDOR platform. Thanks to such an approach new analytics can be defined (either by the technology providers of the project or by the pilots itself, even after the project end), the deployment effort is minimized and it is also possible to protect the source code. All in all, evolution, availability and flexibility of the E-CORRIDOR framework are ensured to accommodate additional requirements and needs identified by the project pilots, even after the termination of the project itself.

Each component to be included in the analytics toolbox needs to comply with and expose a very simple REST-based [3] API (Application Programming Interface) constituted by:

- *START*: to run the analytics over the set of data prepared in the *virtual data lake* (see Figure 1 – component 2) by the ISI
- *STOP*: to gracefully interrupt the execution of the analytics and restore its status
- *KILL*: to abruptly interrupt the ongoing data analysis upon request of the IAI API e.g., in case a policy violation is detected
- *END*: to notify the IAI subsystem of the correct completion of the data analysis and of the availability of the results in the ISI.

Such an interface could be described through a standard and language-agnostic specification (e.g., OpenAPI [4]).

The E-CORRIDOR framework can even accommodate data analytics developed before the project or as off-the-shelf components. In such a case, two possible approaches can be followed. To make it compliant with the controlled data sharing features of the framework a wrapper can be defined, or the analytics can be included in the set of the *legacy analytics engine* (see Figure 1 – component 3). At the time of writing this deliverable at M12, despite the framework support this legacy mode, there is no component identified for being included in such a way.

1.3. Interaction with the E-CORRIDOR Framework

Other than individually calling the analytics in the toolbox, it is possible to compose them in *workflow*s, thanks to the *analytics orchestrator* (see Figure 1 – component 4) provided by the IAI of the E-CORRIDOR framework. Such a composition is constituted by parallel and/or series operations of the original data analytics with the purpose of performing more complex analysis where the output of one component can be combined with the one of others and further analyzed by the subsequent component in the flow.

These workflows can be specified at development or deployment time, are saved in the IAI and are later seen as novel analytics. Thanks to this approach the new analytics are ready to be used by the framework and the user and as a matter of fact transparently perceived as a single service without any additional complexity for the technology providers.

Workflows are specified through a simple Domain Specific Language (DSL) [5] written as a configuration files (e.g., in YAML format [6]) or as short code fragments written in a high level programming language.

Disregarding if the analytics is a simple one in the toolbox or the result of an orchestration, a few basic steps are involved in the functioning of the data analysis in the E-CORRIDOR framework:

1. on the call of an analytics a virtual data lake is created by the Buffer Manager available in the ISI subsystem;
2. the analytics access to the data lake (e.g., through references to the Hadoop Distributed File System (HDFS) [7]) and performs its computation;
3. the data in output are saved as Data Protected Objects (DPOs) through the ISI API;
4. an END message is sent to the ISI to notify that the analysis is complete and the results are ready.

Components in the E-CORRIDOR framework interact through the communications subsystem (see Deliverable D5.2 – "First version of the E-CORRIDOR architecture") and are managed as RESTful [3] services.

For a more detailed description of the E-CORRIDOR architecture and of the ISI and IAI please refer respectively to Deliverable D5.2 and D6.1 ("Sharing and Analytics Infrastructure Architecture).

1.4. Naming Convention for the Analytics and their Requirements

Data analytics in the toolbox and their requirements follow a naming convention similar to the one proposed in D5.1 and here briefly reported for the sake of completeness. Each component is referred as:

E-CORRIDOR-IAI-[Id]

to specify that the components belongs to the analytics infrastructure. The *[id]* is an acronym characterizing the data analytics function.

The sequential number appended at the end refers to the requirements of the component.

1.5. List of Components in the Analytics Toolbox

At the time of writing this deliverable a number of data analytics have been identified by considering requirements and needs of the project pilots. In the following, the analytics components defined at the time of writing this deliverable (M12) are listed grouped according to their purpose:

- Data analytics for driver and passenger identification: sensor data collected from cars, environmental and personal devices are processed by machine learning and artificial intelligence techniques to create models suitable for identification (see Section 2)
    - Secure Routine for driver identification – Driver DNA (E-CORRIDOR-IAI-SR) - Section 2.1
    - Passenger location and flow optimization (E-CORRIDOR-IAI-PL) – Section 2.2

    - Passenger: Identification, Behavior, Context (E-CORRIDOR-IAI-PBI) – Section 2.3
    - Gait analysis – passenger authentication (E-CORRIDOR-IAI-GA) – Section 2.4
    - Face recognition – passenger authentication (E-CORRIDOR-IAI-FR) – Section 2.5
    - Activity recognition – passenger authentication (E-CORRIDOR-IAI-AR) – Section 2.6
- Privacy preserving itinerary planning: by exploiting users' interest and preferences, the best multi-modal travel itineraries are inferred, predicted and self-adapted during the journey according to contextual changes (see Section 3)
    - CO2-aware Trip Planning (E-CORRIDOR-IAI-MMIP) – Section 3.1
- Privacy preserving (Security) analytics: privacy-preserving analytics based on fully homomorphic encryption and secure multi-party computation are used to perform security checks on user shared data (see Section 4)
    - OpenAPI for Fully Homomorphic Encryption (FHE) (E-CORRIDOR-IAI-FHEC) – Section 4.1
    - Secure Multiparty-computation for Routine based authentication - Private Secure Routine (E-CORRIDOR-IAI-MPCSR) – Section 4.2
- Carbon foot print analytics: data collected in real time are used to infer by approximation the environmental impact of multi-modal journeys (see Section 5)
    - CO2 Analytics (E-CORRIDOR-IAI-CFA) – Section 5.1
- Intrusion detection technologies (IDS): machine learning techniques are used to perform anomaly-based detections and to enforce cyber-security on all the transportation entities (see Section 6)
    - Automotive Intrusion Detection (E-CORRIDOR-IAI-CANIDS) – Section 6.1
    - Fully Homomorphic Encryption-based intrusion detection (E-CORRIDOR-IAI-FHEIDS) – Section 6.2
    - Intrusion Protection System (IPS) – EARNEST (E-CORRIDOR-IAI-CANIPS) – Section 6.3
- Pilot specific analytics: are components that, even if totally compliant with the E-CORRIDOR framework and included in the toolbox, will run exclusively on the pilot premises mainly for design decision and being tailored for a specific application (see Section 7).

Note to the reader: As the deliverable aims at describing all the data analytics techniques used in the E-CORRIDOR project, there are a multitude of different technologies in a single document that often require disparate technical knowledge. To help the readers in navigating the document, in the list above, a reference to the corresponding section where the component is described has been provided. Each description is self-contained (including contribution to the platform requirements, application to pilots and synergies) so that the reader can point to the

data analytics components more relevant for her own interests without reading all the previous descriptions.

It is worth to remark that the above set of analytics have been selected by taking into account the input received by the project pilots at the time of writing this document. But thanks to the plugin approach new analytics may be added in the toolbox to accommodate any potential additional need raised during the execution of the E-CORRIDOR project. The same list of components available in the analytics toolbox, along with their logical grouping, is depicted in Figure 2.

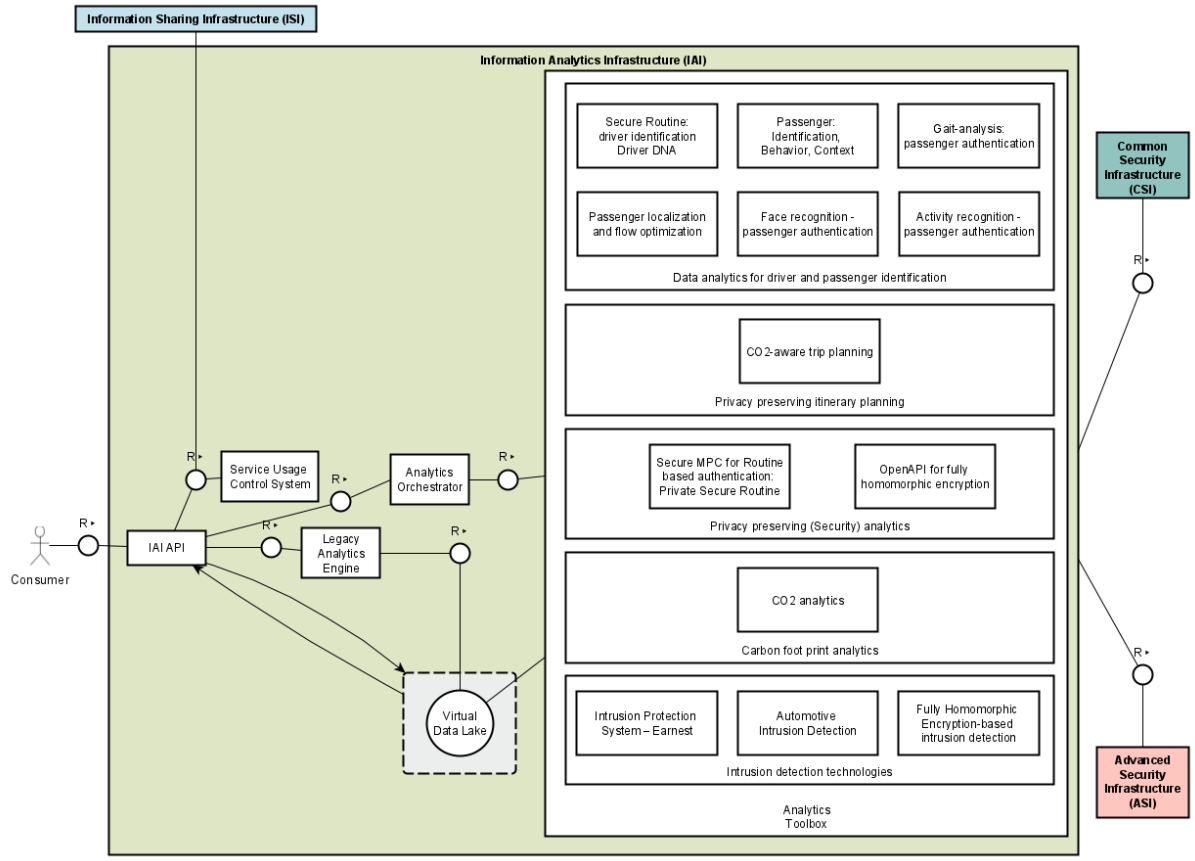

Figure 2 Analytics Toolbox in the IAI subsystem of the E-CORRIDOR framework.

The analytics in the toolbox at M12 have a Technology Readiness Level (TRL) [8] spanning from level 3 to 4 (i.e., the research shows that the tool is feasible). Thanks to the support of the pilots and the evaluation of the components in their environments, throughout the execution of the project is expected to maturate the components up to TRL 6 or 7 (i.e., the tool demonstrated its capability in a realistic environment) and therefore support the achievement of Objective 4. Finally, being the analytics defined by taking into account the pilots requirements, the

challenges of their domains and their specific needs, the same will be used to deliver pilot products (Objective 6) and potentially even ease their adoption (Objective 7).

All the tools in the toolbox are generally meant to be executed in a hybrid edge-cloud fashion, even if specific restriction could be imposed by the technology providers (e.g., in case of specific hardware requirements) or by the application.

2. Data Analytics for Driver and Passenger Identification – Task 7.1

The components related to this task analyze sensor data to perform driver and passenger identification. Sensor data are collected from cars in case of the S2C pilot (e.g., OBD readings, GPS, CAN bus messages) or environmental and personal devices in case of the AT pilot (e.g., cameras, Bluetooth beacons, wearable and smartphone sensors). Thanks to machine learning and artificial intelligence algorithms these data are analyzed to create models for driver and passenger used for identification. The same models can also be exploited for behavioral and driving style analysis other than for authentication purposes.

2.1. Secure routine for driver identification - Driver DNA [E-CORRIDOR-IAI-SR]

Even though the introduction of ICT in transportation systems leads to several advantages in terms of efficiency of transport, mobility, traffic management, and in improved interfaces between different transport modes, it also brings some drawbacks in terms of increased security challenges, also related to human behavior. For this reason, in the last decades, attempts to characterize drivers' behavior have been mostly targeted towards risk assessment and, more recently, to the training of machine learning software for autonomous driving.

Driver behavioral characterization can be used to build a general reputation profile that can help to create innovative, reputation-aware automotive services. As a first step towards realizing this vision, we present guidelines for the design of a privacy preserving way to collect information generated from vehicles sensors and the environment, and to compose such collected information into driver reputation profiles. In turn, these profiles are exchanged in a privacy preserving way within the infrastructure to realize reputation-aware automotive services, a sample of which are described in the following. As a fundamental component of the infrastructure, we show that: i) multi-dimensional reputation profiles can be formed building upon the recently introduced notion of driver DNA [9]; ii) multi-dimensional comparison of profiles can be achieved by means of a reputation lattice rooted in the notion of algebraic c-semiring; and iii) a secure two-party mechanism can be used to provide services to drivers on the basis of their reputation and/or DNA's parameters.

2.1.1. State of the Art

In the last few years, interest about the characterization of driver behavior according to information collected from the vehicle has consistently increased. However, to the best of our knowledge, none of the existing work attempts to link driver behavior to the notion of reputation and trust.

One of the early works in this field is presented in [10], where the authors proposed a traffic simulation model incorporating assumptions about what a safe drivers' behavior should be. The main outcome of the paper is the comparison between results obtained in the simulation and the real world.

Other recent works [11], [12], present approaches to identify reckless drivers based on a combination of speed and acceleration. Both measures are retrieved from different ICT systems present in the vehicle itself. In [12], the information was retrieved from SD Card and GPS on vehicle.

In [13], the driver is considered as part of the vehicle system (driver-in-the-loop), more specifically as the control unit of the entire system. In this way, the authors described three methods to identify driver behavior as a comparison with the actual and the expected behavior of the system by considering different aspects of the drive-in-the-loop vehicle system.

Works about how to link the driver behavior with traffic accidents, safety on roadside network, and possible rewarding are mostly related to the insurance world. For instance, reference [14] is about the risk of reckless drivers and how insurance reward can depend on the driver behavior. Adapting insurance fee to driver behavior is promoted as a method to incentivize drivers to drive more carefully and reduce accidents.

To our best knowledge, the idea of characterizing driver's behavior with the final aim of computing a comprehensive driver's reputation profile and to realize reputation-aware vehicular services is a novelty of this component and presented for the first time in [15].

About reputation-aware vehicle service, several services for Intelligent Transportation Systems (ITS) have been introduced in the literature. Following the standardization work of European Telecommunications Standards Institute (ETSI), ITS applications (or service) have been categorized in a number of classes. While their requirements and operational constraints have been defined in ETSI, security specifications are not fully defined and mostly left to the single developers. For instance, secure and privacy aware versions of two representative classes of ITS applications are Driver Assistance – Road Hazard Warning, and Community Services. In case of road hazard warning, there is ample literature that studies under what conditions the communication network (V2V and V2I communication) is able to provide the adequate level of responsiveness necessary to enable early hazard detection [16]. Since security and privacy requirements as mandated by the proposed architecture will introduce significant communication/computational overhead, there is a need of carefully analyzing and testing the interplay between security level, communication performance, and achieved effectiveness in providing secure and early warning to the drivers.

2.1.2. Proposed Approach/Technology

The notion of Driver DNA [9] has been proposed to concisely represent a driver's driving style starting from car-collected data analysis, integration with road and weather information, and comparison with peer drivers. The Driver DNA has been firstly proposed in [15] and it is made of four parameters: braking (b), turning (t), speeding (s), and RPM (rpm) (revolutions per minute). These four parameters are not directly comparable. Each parameter is measured with a rank ranging between 0 (lowest score) and 5 (highest score). The first parameter (braking intensity) is used to quantify a driver's aggressiveness, the second (steering wheel angle) is used to quantify comfort in driving, the third parameter (driving above speed limit), which is also combined with weather information, is directly related to accident risk, while the fourth parameter (engine RPM) is used, when compared with values obtained by peer drivers, as a proxy of a driver's fuel efficiency.

Following [9], we represent the profile of each driver as a tuple of four elements (b_i,t_i,s_i,rpm_i), with $b_i,t_i,s_i,rpm_i \in [0, 5]$, one for each parameter we are going to consider to identify the driver's DNA. Using the profile, we associate to each driver a reputation value.

In fact, the authors of [9] suggests graphically representing a driver's driving style as a radar graph of the four dimensions, where a relatively larger area of the radar graph indicates a relatively better driver. Thus, the Driver Reputation score (RD_i) is represented by the internal area identified by the radar graph derived by the four parameters of the driver's DNA.

Hence, each driver in the E-CORRIDOR architecture can be characterized by a multi-dimensional reputation profile, which should be considered as a valuable and private information to the driver. Reputation profiles of drivers become a sort of passport in the E-CORRIDOR framework. Thus, they can be exchanged in a secure and private way with

surrounding vehicles and roadside infrastructure to realize innovative reputation-aware vehicular services

In fact, vehicles in E-CORRIDOR can ask for services, getting different quality and or prices depending on their driver's reputation profile. Typically, we can assume that to obtain, e.g., a special discount on a service, a driver must provide her profile to be compared with an access threshold used by the service provider. This comparison function hits the driver's privacy since the service provider will be able to know the entire profile in case of full profile disclosure, or at least a single parameter in the reputation profile.

To protect the privacy of the drivers, we implemented the comparison function in a privacy-preserving manner that make use of the Secure Two Party Computation (2PC) technique CBMC-GC v1.0 [17] that allows drivers to discover whether they meet the conditions for obtaining a certain service level without disclosing their profile.

Examples of innovative "reputation-aware" services enabled by this component in the E-CORRIDOR framework are described:

Reputation-aware fuel cost. Currently, fuel cost is decided at the level of the single gas station, and it is applied independently of the driver's attitude to save or waste fuel while driving. In an effort to incentivize fuel-efficient driving style, one might think of a scenario where fuel cost is personalized to reflect a driver's fuel efficiency. When entering a gas station, the vehicle onboard software sends driver's reputation information – in this specific case, both her reputation score and her fuel efficiency score – to the fog node installed at the gas station. After proper authentication, the driver will be offered a personalized fuel price: a relatively lower price for drivers with relatively higher reputation, and vice versa.

Reputation-aware tolling. Similarly to the case of fuel price, also access to road infrastructure is currently oblivious to driving style, and is typically done based on the type of vehicle. However, a driver with a relatively higher risk profile (e.g., more aggressive, or speeding more frequently) might pose a relatively higher prospect cost to the infrastructure manager than a relatively more cautious driver, due to the higher risk of incurring accidents, damage road components, etc. One can then envision a scenario in which the price to access road infrastructure (highways, bridges, etc.) is personalized based on a driver's reputation profile. Similarly to the gas station scenario, the vehicle onboard software shares driver's reputation information with the fog node interfacing with the tolling system, and a driver is charged a variable amount that reflects her accident and damage risk profile.

2.1.3. Data Format Requirement

There is not a required format of data. It depends on the in vehicle and environmental sensors. The output is a Boolean flag that answers to the question about the goodness of the driver reputation. E.g., the access to the services can be granted only to the drivers having a reputation above a predefined threshold.

2.1.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-SR-01	MUST	Data used to establish the driver DNA may require to be obfuscated or anonymized, e.g., data coming from vehicles that may provide information on drivers	• E-CORRIDOR-DM-01 • E-CORRIDOR-Sec-RC-01
E-CORRIDOR - IAI-SR-02	MUST	Driver DNA analytics can be run at the edge.	• E-CORRIDOR Ope-02
E-CORRIDOR - IAI-SR-03	SHOULD	The In-Vehicle Infotainment (IVI) or Electronic Control Units (ECUs) may be used for collecting GPS and driving behaviour data.	• E-CORRIDOR-Tst-S2C-01 • E-CORRIDOR-Tst-S2C-02

2.1.5. Application to Pilots

Pilot	S2C, ISAC
Reference to Use cases or User stories	• S2C-US-08: Driving behavior recognition • ISAC-US-01: Public cyber-threat information collection
Brief description of the Use cases or User stories	The above use cases refer to the possibility of providing customized services to the user. The reputation-based approach we propose aims at analyzing data collected by sensors belonging to the framework (vehicles, environment, etc.) in order to provide useful and customized services depending on the Pilot use cases.
Match of the proposed approach/technology with the USs/UCs	This analytics will help the stakeholders to achieve incentives depending on “reputation-aware” services and based on drivers style. Currently this component targets the automotive infrastructure but, in the future in principle, it can extended to build reputation and provide incentives even to the passengers of the AT pilot.

2.1.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.3
Title/brief description of the task	The above task refer to Privacy aware interest-based service sharing.
Description of the potential synergy with risks and opportunities	According to the driver's reputation and the application of Secure Two Party Computation, it is possible to provide services to users in such a way that drivers' data are shared in a privacy preserving way.
Dependencies on other components	None

2.2. Passenger location and flow optimization [E-CORRIDOR-IAI-PL]

Indoor localization is a mature research area which gained a lot of traction in the recent years due to the availability of sensors supporting localization applications on smartphone devices. *Outdoor localization* services are based on accurate Global Navigation Satellite System (GNSS) sensors and now counts as many as 6.4 billion enabled devices worldwide [18]. The same high level of accuracy that these sensors provide in outside environments is not achievable in indoor settings due to signal attenuation, interferences and to loss of continuity and reliability of the service [19]. For these reasons, *indoor localization* requires a different approach based on different type of sensors. The idea of this component would be to utilize the Received Signal Strength Indicator (RSSI) of the Bluetooth Low Energy (BLE) device to estimate the position of the people moving indoors. We opted to use BLE signals over Wi-Fi and Micro-Electro-Mechanical System (MEMS) accelerometers and gyroscopes signals due to its intrinsic characteristics as well as its affordability and availability on common smartphones and wearable devices. BLE signals have low range of effect (approximately 10-15 meters) which allow us to both locate people in a small area within the reach of the point of interests as well as to keep the signal communications low and localized in crowded areas, unlike Wi-Fi signals which have higher reach and would considerably increase the traffic generated to locate all the people in an area. Moreover, BLE-based localization would be agnostic to the different characteristics of how the person moves within the environment (e.g., type of walk, length of steps), and makes the inference of its position easier compared to the analysis and tuning that would be required if MEMS sensors would be deployed.

With respect to the E-CORRIDOR project pilots, it would represent the passenger movements within the airport and train station premises. BLE devices are gaining more relevance due to their wide availability, both in environmental (e.g., Internet of Things, IoT) and personal (including wearable) devices. Estimating the passenger location with a high accuracy will enable further possibilities, such as passenger flow optimization, customer assistance and guidance in an unknown wide environment and tailored services. Furthermore, the location of the passenger can also be exploited to increase the accuracy of her authentication by expanding the set of contextual information.

2.2.1. State of the Art

BLE technology has become widely available and affordable and can be used for many indoor activities, such as location estimation. Several research activities have explored indoor localization and user movement by leveraging the data gathered by MEMS accelerometers and gyroscopes, as presented by [20], [21], [22]. Despite a comparable and low cost for both BLE and MEMS devices, less effort has been devoted to BLE solutions and datasets exploiting BLE signals for indoor localization are rarer to find.

The work done in [23] tries to overcome the lack of proper datasets and benchmarks. Their dataset has been generated in an indoor environment using BLE anchors (also referred as beacons) and smartphones replicating different scenarios with multiple actors. The scenarios have also considered the use of different transmission powers to send messages between devices as well as having two opposite types of communications (beacon to device and vice-versa). Despite its limited size in terms of scenarios, this dataset represents a valuable source for studies aiming at performing BLE-based localization and can be leveraged to push the research and innovation for indoor localization systems based on the BLE technology.

BLE devices rely on RSSI values, which represent the intensity of the signal received by the device. These values can be used to estimate the position of the signal sender, as shown in [24] and [25], where the signal is used to train a probabilistic model based on the Dempster-Shafer theory to estimate the sender's position. The estimation relies heavily on the external knowledge regarding the environment in which the user device is moving and also on the number of beacons used.

A further interesting use of the BLE technology is the possibility to trace the interactions that different devices can have with each other. BLE devices have low range of action, a physical constraint given by the BLE technology. This limit can be exploited to track the proximity of two devices to each other with lower interferences over other technologies (e.g., WiFi). This characteristic has been proven useful and has been applied to help in tracking the contacts between different people and therefore to track the potential spread of the COVID-19 infections as presented in [26]. This work considers not only the tracking capabilities but also the possibility to do so while keeping in consideration the privacy of each user. The latter is one of the key features for the E-CORRIDOR framework and use cases, and an aspect deemed particularly relevant by the project pilots.

2.2.2. Proposed Approach/Technology

The passenger localization and flow optimization tool aims to: (i) estimate the position of a passenger when she is moving in an indoor environment (i.e., train station and airport areas), (ii) gather all the estimated passengers' locations and use them to depict a picture of the passenger flow within the environment and, potentially, use this information to optimize critical crowd bottlenecks (by either redirecting the flow in less crowded areas or by increasing the personnel in such zones) and, (iii) guide the passengers throughout their journey with personalized privacy-aware messages to direct them to their points of interest.

The first step that has to be achieved is having an accurate estimation of the position of the passenger inside the environment. We plan to use the Bluetooth technology for this task. We opted to only use BLE RSSI values over a mixture of RSSI and MEMS accelerometer/gyroscope data so that both the computation required to track the passenger as well as the amount of information shared between the stakeholders would be kept to a minimum. We assume that the indoor premises are equipped with several stationary BLE beacons (or anchors) that will communicate with the passengers' BLE devices. The RSSI values

exchanged between the devices will be used to estimate the passenger position: the distance between the beacon and the passenger device is assumed to be directly proportional to the distance, i.e., the closer the device is to the beacon, the more powerful the signal exchanged will be, the further away the device is to the beacon, the less powerful the signal power will be. To enhance the estimation accuracy, it would be ideal to have the whole premises fully covered by overlapping beacons, so that the passenger will always be under the coverage of at least a couple of beacons. An example of a passenger's movements and the BLE beacons to which her device can communicate with is shown in Figure 3.

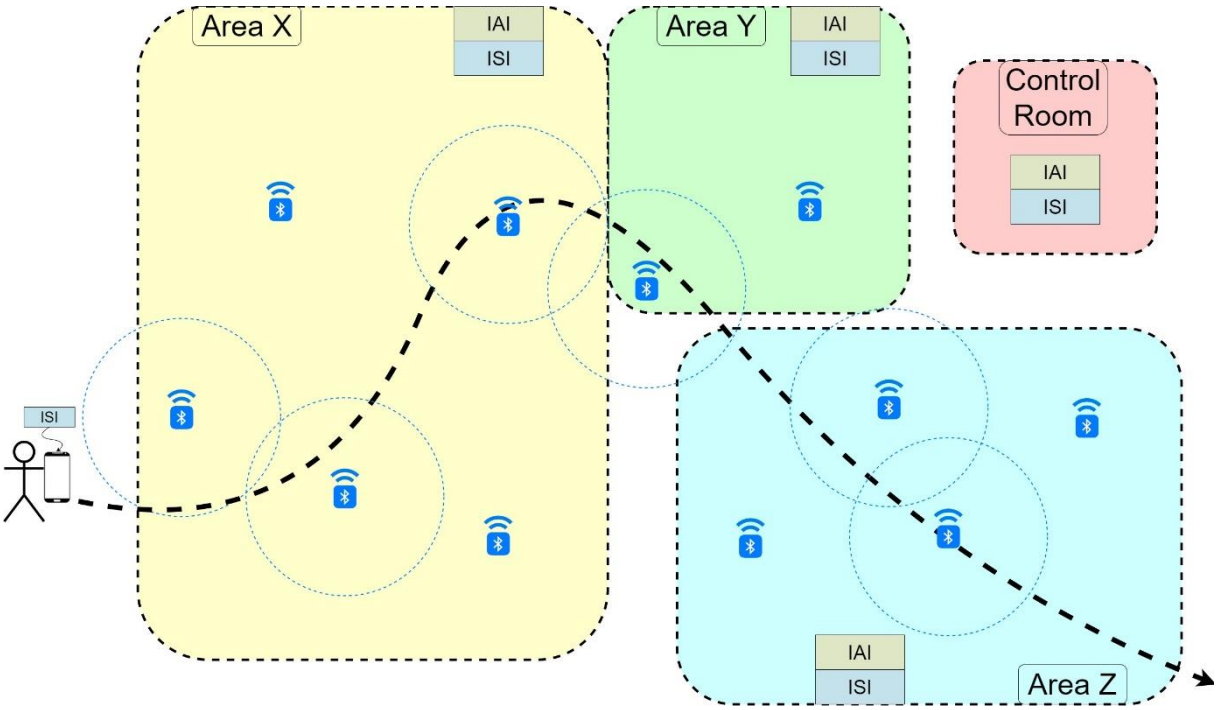

Figure 3: Passenger movement detected by different BLE beacons and the deployed E-CORRIDOR framework.

An indoor environment can be subdivided in *virtual zones*, e.g., according to the displacement of environmental sensors (cameras, lidars, BLE beacons, etc.). Each virtual zone in which the premises are divided (in the Airport-Train pilot it could represent terminal, lounge, etc.) can have an instance of the E-CORRIDOR framework at the edge (running the whole framework with the IAI and ISI subsystems or a lightweight version of it). Thanks to the controlled and privacy-aware data sharing capabilities of the E-CORRIDOR framework, by collecting anonymized information on number and position of the devices in the area, a holistic view of the whole environment can be generated. The latter would be useful for managing critical infrastructures such as the ones represented by airport and train station. The managers of the infrastructure can thus access this information from the control room (where the master local instance of the E-CORRIDOR framework is supposed to be located) to generate an approximate map of the environment, how the flow is moving, which are the bottlenecks that require attention and apply mitigation or recovery strategies.

Furthermore, by deploying the beacons in specific points of interest of the infrastructure, such as security screening and check-in kiosks, the estimated travel and sojourn times can be inferred and more thorough analysis can be built on top of this information. Similar pilot projects have been proposed and developed in Milano Malpensa airport (Italy) by [27]. This would be

enhanced by the privacy-aware capabilities and collaborative sharing and access to information provided by the E-CORRIDOR framework.

The localization system can even work in a setting where beacons transmit information about their position in the virtual zone. Then, a lightweight E-CORRIDOR framework, including the IAI subsystem and the needed analytics in the toolbox, is installed in the passenger's smartphone, infers the passenger position and interacts with the E-CORRIDOR node deployed in the zone to receive the above described services.

The last purpose of this tool would be to guide passengers throughout their main point of interest in the indoor environment. This is enabled by the E-CORRIDOR framework which could provide information regarding the passenger flight details and guide her from the correct check-in kiosk until her departure gate. To achieve such a *turn-by-turn navigation*, we plan to adopt the Eddystone protocol [28] and specialize it to send dedicated privacy-aware messages to each passenger. The edge ISI will have access to the needed information of the passenger with respect to her travel. This can be combined with the BLE localization component to help design her navigation throughout the environment. We envision that the only passenger related information shared in these navigation messages (from the E-CORRIDOR framework instances running in the virtual zone and on the passenger's device) will be an identifier based on the Bluetooth device that she is using. Furthermore, the Eddystone protocol has a dedicated type of encrypted messages to further enhance the privacy of the exchanged messages.

2.2.3. Data Format Requirement

The data expected would mainly be the RSSI values of each device with corresponding identifiers of sender and receiver, and a timestamp. An example of the dataset presented in [23] is extracted and shown below.

Timestamp	Sender Identifier	Receiver Identifier	RSSI (signal strength)
1540748414903	1070	1	-69
1540748414903	1064	1063	-72
1540748414912	1063	1062	-79
1540748414912	1064	1062	-79
1540748414914	1062	1069	-81

The output would be a representation of the estimated position of the devices within reach of BLE anchors within the airport/train station.

2.2.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil D7.1 Requirement(s)
E-CORRIDOR-IAI-PL-01	MUST	The accuracy and effectiveness of passengers' localization and specific guidance messages depend on the DSA specified by each passenger, on the attributes of the latter and on contextual/environmental properties.	• E-CORRIDOR-DS-06 • E-CORRIDOR-DS-07 • E-CORRIDOR-DS-16 • E-CORRIDOR-DS-17 • E-CORRIDOR-DS-24
E-CORRIDOR-IAI-PL-02	MUST	The passengers' flow estimation will rely on the information gathered by all the passengers that allowed their data to be analyzed.	• E-CORRIDOR-DS-27 • E-CORRIDOR-DA-05 • E-CORRIDOR-DA-07
E-CORRIDOR-IAI-PL-03	MUST	The guidance messages directed to specific passengers will be generated considering the privacy of the passenger.	• E-CORRIDOR-DA-10 • E-CORRIDOR-Ope-05
E-CORRIDOR-IAI-PL-04	SHOULD	The estimated position of each passenger should be used to enhance her authentication.	• E-CORRIDOR-Use-02
E-CORRIDOR-IAI-PL-05	MUST	The inferred passenger location is transmitted and stored (only for the needed time) in a privacy-aware and secure manner.	• E-CORRIDOR-DS-10 • E-CORRIDOR-DA-11 • E-CORRIDOR-Sec-IS-02
E-CORRIDOR-IAI-PL-06	SHOULD	The analytics to locate the passenger can run either on the personal device (i.e., at the edge) or on the cloud.	• E-CORRIDOR-Ope-01 • E-CORRIDOR-Ope-02

2.2.5. Application to Pilots

Pilot	Airport-Train pilot
Reference to Use cases or User stories	• AT-US-01: Passenger Management and Operations • AT-US-03: Distributed and Combined Context Analysis in Sensor Network • AT-UC-01: PRM Passenger Assistance and Authorization • AT-UC-04: Privacy-preserving Passenger Monitoring • AT-UC-09: Sharing of Service Access Data • AT-UC-11: Notification of Service Disruption • AT-UC-12: Passenger Flow and Prediction • AT-UC-14: Notification on PRM Passengers' Location
Brief description of the Use cases or User stories	The above use cases and user stories refer to passenger localization, monitoring and contextual services. Moreover, the owners of the infrastructure (i.e., airport and train station) can optimize their services by sharing data in a privacy-aware manner.
Match of the proposed approach/technology with the USs/UCs	From the point of view of the passenger, location-based information can be used to achieve a stronger authentication mechanisms and to receive better and tailored services (e.g., turn by turn navigation). The same information can be used by the airport and train station to improve operational and provide additional services.

2.2.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.1 • T8.2 • T8.3
Title/brief description of the task	The above tasks refer to privacy aware interest-based service sharing and the passenger's contextual authentication.
Description of the potential synergy with risks and opportunities	The data gathered from the passenger have to be shared within the E-CORRIDOR framework respecting the privacy policies set by the passenger itself. The same information can also be used to expand the contextual behavioral authentication of the passenger.
Dependencies on other components	None

2.3. Passenger: Identification, Behavior, Context [E-CORRIDOR-IAI-PBI]

With the general increase of the number of passengers in airports and train stations experienced in the recent years, the number of deployed cameras has also been increased to maintain monitoring activities and behavior characterization of passengers. In 2019, Paris Charles de Gaulle Airport received 76 million passengers. Accordingly, there has also been an increase at the level of workload of video operators to analyze and understand video content. Automated analysis of large amounts of data is needed to process the data in real time and significantly enhance passengers monitoring, identification and analysis.

Automatic monitoring and identification in crowded areas will afford continuous monitoring and behavior analysis without relying on constant human interaction. Monitoring numerous people via multiple cameras is a challenging task, especially in complex and crowded areas such as airports and train stations with frequent occlusions and interaction between groups of individuals.

In this context, we present an analytics to monitor, identify and characterize the environment surrounding the passenger.

2.3.1. State of the Art

Object monitoring is an important research area in Artificial Intelligence (AI) with a wide range of applications, such as visual surveillance systems. Person detection using visual surveillance system is relying on manual methods of identifying unusual activities. However, it has limited capabilities and it is not very effective [29].

Throughout the last decade, to monitor and capture person activities, the visual object tracking technologies have achieved significant progress, especially when deep learning has been applied, making the person monitoring a breakthrough. Convolutional Neural Networks (CNN) have been largely adopted in learning complex systems where a single model includes all the intermediate necessary processing steps. CNN are therefore considered a reliable choice for end-to-end learning and have been applied for image representations where images are classified and directly mapped to the target labels (e.g., describing the objects present in the scene). Convolutional architectures have been used for solving supervised learning problems and assessing images for different applications. [30] used CNNs for predicting the location and extracting features of an individual user for tracking. [31] predicts trajectories of people based on their last positions using Long-Short Term Memory (LSTM) networks.

In presence of multiple domains, to increase the effectiveness of the system, person re-identification is often considered as a desirable feature. Person re-identification is defined as the task of associating the presence of the same person in different places at different times as detected by multiple environmental cameras. A comprehensive and recent review on the multi-object tracking methods has been presented by [32]. In computer vision, person re-identification techniques are applicable to both static images and videos [33]. [34] used deep learning approach for re-identification. They focus on finding an improved network architecture, an effective set of features and defining a similarity function for comparing those features. [35] used group context by proposing ratio-occurrence descriptors to capture groups of people. [36] adopted spatio-temporal relationships via cameras for person re-identification.

2.3.2. Proposed Approach/Technology

In this analytics, a deep learning based framework will be adopted to perform passenger and baggage detection and monitoring. Figure 4 shows a diagram of the proposed architecture of the component.

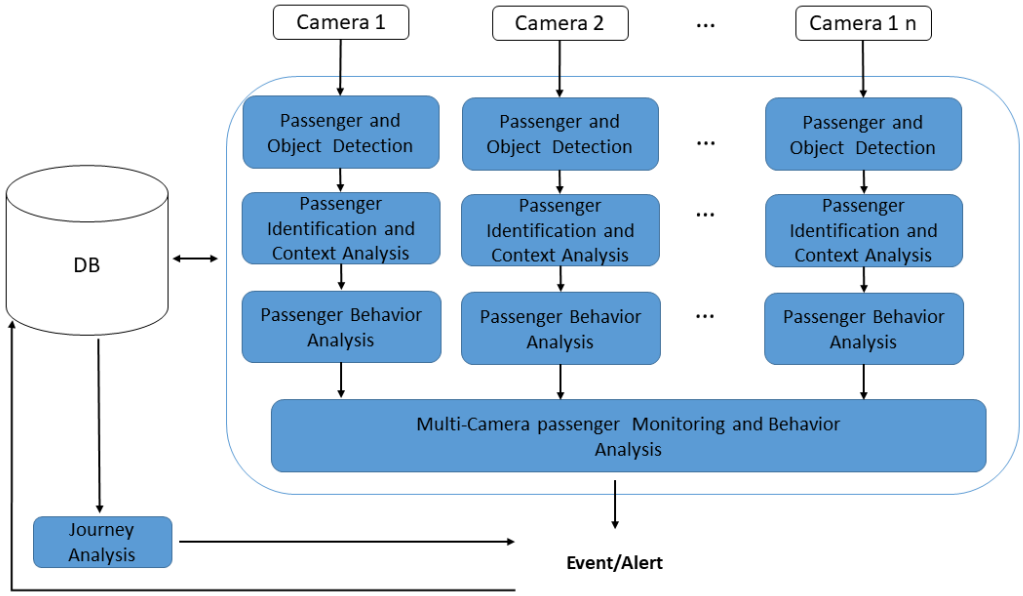

Figure 4 Architecture of the Passenger: Identification, Behaviour, Context component

Every camera will be linked to three modular components. The first module takes in input a video stream and performs passenger detection. Detected passengers will be marked through bounding boxes and a temporal identifier (id) will be assigned to every passenger in the main entrance of the train station or the airport.

The role of the second module is to identify every passenger. The temporal id will be replaced by the real identifier of the passenger as collected by the self-service kiosk and made available through the ISI and appropriate DSA. The environment surrounding the passenger will be used to enhance the identification during the detection phase and new classes representing the environment surrounding the passenger such as backpack, handbag, suitcase and wheelchair will be linked to the main model (representing the passenger).
Afterwards, the third module takes every identified passenger and keeps monitoring for identifying their position, behavior and movement in the premises of the transportation infrastructure (either train station or airport).
The results of the passengers monitoring will be logged into a database and stored for a defined time window. The journey analysis module can then perform additional analysis of the passenger behavior based on the collected data to extract further knowledge (e.g., to infer the experience of the passengers in a given terminal). This process can lead to support, as well as abnormal behavior detection.
The multi-camera passenger monitoring and behavior analysis module will be used to match the scenario constituted by multiple cameras. It allows for passenger monitoring across multiple cameras, whether or not the cameras have overlapping field of-views.
In order to maintain the balance of accuracy and real-time monitoring, YOLO v4 [37] as an object detection algorithm alongside with the Deepsort [38] as a tracking approach will be utilized.

The proposed component will allow the analysis of the passenger behaviours and their activities for high level surveillance tasks e.g., event and activity detection, crowd analysis or groups movement as well as infer group of passengers travelling together (that may be linked to the same Passenger Name Reservation, PNR). Thanks to these capabilities, the proposed analytics component can have several safety and security applications: suspicious behaviour prediction, criminal tracking, passenger tracking in a defined area, searching for lost children, monitoring of passengers with limited mobility to promptly identify the need for additional special services, etc.

2.3.3. Data Format Requirement

The input data can be described by a JSON file representing the set of cameras as a graph where nodes define the characteristics of the cameras and edges represent the connection between cameras if they corresponding covered areas are directly linked. An example of two connected cameras is described in the JSON file as follow:

```
{
  "directory":../data/videos/",
  "adjacency":{
    "Cam1":["1"],
    "Cam2":["2"]
  },
  "nodes":{
    "Cam1":{
      "file": "top.mp4",
      "Char":{}
    },
    "Cam2":                                                    {
      "file": "bot.mp4",
      "Char":{}
    }
  },
  "edges":{
    "1":{
      "from":"Cam1",
      "to":"Cam2",
      "data":{}
    },
     "2":{
      "from":"Cam2",
      "to":"Cam1",
      "data":{}
    }
}
```

2.3.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-PBI-001	SHOULD	The collected features of every passenger should be used to enhance the authentication process.	E-CORRIDOR-Use-02
E-CORRIDOR-IAI-PBI-002	MUST	The cloud service, after finishing the travel journey must delete the subject information.	E-CORRIDOR-DS-10
E-CORRIDOR-IAI-PBI-003	MUST	The collected passenger features are transmitted and stored in a privacy-aware and secure manner.	E-CORRIDOR-Sec-IS-02

2.3.5. Application to Pilots

Pilot	AT pilot
Reference to Use cases or User stories	• AT-US-03: Distributed Tracking Analysis in Sensor Network • AT-US-05: End to End Safe-Contact/Contactless Journey • AT-US-07: Document-free Secure Multimodal Travel Credential
Brief description of the Use cases or User stories	The above use cases refer to the possibility of providing a frictionless experience to the user while accessing to the transportation system. .
Match of the proposed approach/technology with the USs/UCs	The contextual and behavioural analysis can be used to support contactless and biometric-based access to the transportation services.

2.3.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.1 • T7.1
Title/brief description of the task	The above tasks refer to Multi-Biometric/Factor Authentication, localization services and activity tracking
Description of the potential synergy with risks and opportunities	Camera based analysis can be used along with other passenger data collected from other sensors and the analysis of other analytics components like gait from smartphone IMU, location based on BLE beacons, and RFID from passport, to perform a seamless and strong authentication through a multi-biometric and multi-factor approach.
Dependencies on other components	Multi-biometric and multi-factor authentication – T8.1

2.4. Gait analysis – passenger authentication [E-CORRIDOR-IAI-GA]

The authentication of a passenger is a fundamental procedure to check the person in every transportation hub, e.g., airport, train station, car station. Due to the long process that each passenger has to undergo before boarding, he must check-in, in some cases, few hours prior to their travel. The introduction of new authentication mechanisms based on biometrics, (e.g., face recognition, iris recognition, and fingerprint), has made it possible to reduce the computation time and increase the recognition accuracy. Despite this, the great diffusion of personal and wearable mobile devices able to collect unobtrusively data related to the user behavior has made more relevant new scenarios based on seamless authentication. With seamless authentication, biometric features such as human gait become a way to control authorized, without actually requiring user interaction. However, this analysis is a challenging task, prone to errors, with the need to dynamically adapt to new conditions and requirements brought by the dynamic change of biometric parameters. Gait recognition, or the measurement of a person's walking pattern, may reach the accuracy of face recognition with the advantage of being less intrusive.

The new biometric security system, based on gait analysis, can be exploited in a constrained path, like the gate check-in any transportation hub, for authentication and authorization purposes (e.g., to verify that the ticket corresponds to the passenger who has performed the check-in).

2.4.1. State of the Art

Advances in wearable technology have added computing capacity and sensors into smartphones, tablets, (smart) watches, but also shoes, clothes, and other wearable items. These enhanced objects act as enablers of pervasive computing, collecting data to provide additional smart services to their users. Several of these smart devices come equipped with built-in accelerometers and gyroscopes, which can be exploited to register the body motion of users useful for seamless authentication. However, most current solutions for sensor-based

authentication are mainly based on active behavioral mechanisms, which require direct user interaction [39].

The evolution of machine learning and its application to inertial data has been largely used also in the security field. Specifically, it is applied to authentication mechanisms based on soft biometrics. Various techniques have been proposed to analyze the behavioral usage of the smartphone through sensory data. In [40] and [41] are proposed authentication mechanisms where additional external devices perform the collecting phase. Specifically, in [40], an algorithm that exploits dedicated external hardware to gather and analyze inertial gait data is proposed. In [41], a multi-class machine learning algorithm is applied to users' identity verification analyzing a different number of activities registered by heterogeneous sensors.

A less intrusive system is implemented in [42]. The author proposed a deep-learning-based active authentication approach that exploits sensors in consumer-grade smartphones to authenticate the user. In [43] the authors use the gait analysis to identify users through a convolutional neural network.

Most of the gait analysis works are focused on a unique type of walking pattern, whereas few works are interested in analyzing different kinds of walking actions. To this end, [44] introduced Hand Movement, Orientation, and Grasp (HMOG), a set of behavioral features to authenticate smartphone users continuously. HMOG features unobtrusively capture subtle micro-movement and orientation dynamics resulting from how a user grasps, holds, and taps on the smartphone when a user is walking or sitting. A machine learning mechanism trained on different walking actions allows to obtain a more robust learning process and, thus, an authentication mechanism that is less invariant to the user movement.

2.4.2. Proposed Approach/Technology

In the E-CORRIDOR framework, a machine learning based mechanism will be defined to implement an unobtrusive authentication system based on gait analysis usable in every transportation hub e.g., to verify that the ticket corresponds to the passenger that had performed the check-in. The system exploits a background smartphone application that collects inertial data from the gyroscope and accelerometer sensors and send them to a machine learning component that recognizes and learns the walking path of the person. The learnt walking features will be stored in a database when the person buys the travel ticket. Later, when the passenger enters in a gate check-in, the smartphone application will send the walking inertial data to the system that will extract the walking features and will compare them with the ones stored in the database. The application of the proposed system is described in Figure 5.

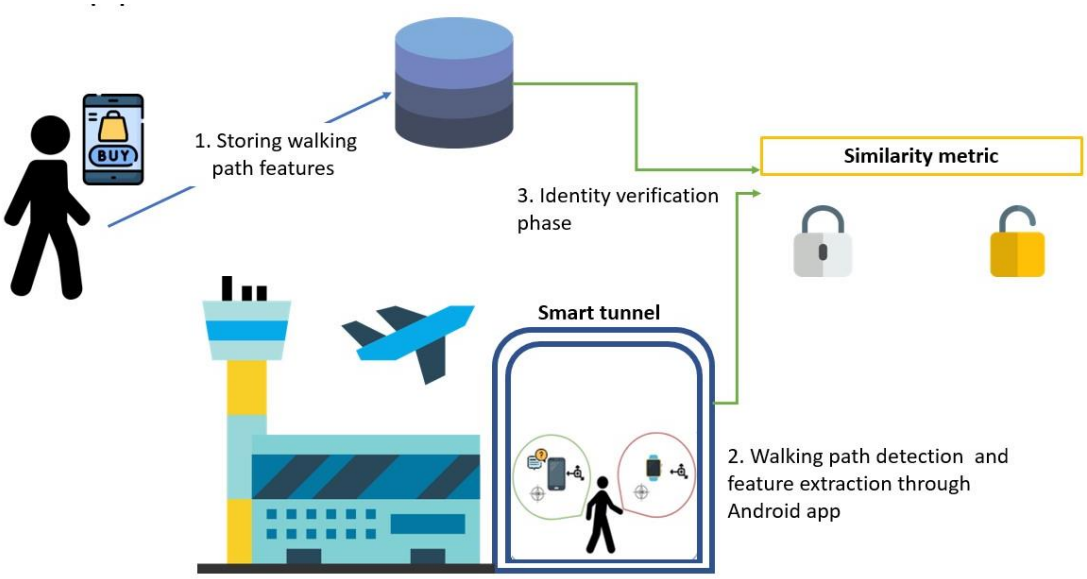

Figure 5 Gait analysis for passenger authentication in the transportation domain.

The core component of the system is a Human Action Recognition (HAR). It is a deep learning network based on Recurrent Neural Network (RNN), able to analyze the inertial signal provided by the preprocessing component and infer the specific movement or action that a person is performing. This component is embedded in the Android application and analyzing the inertial data in a time window, it classifies the user activity as running, walking, walking upstairs, walking downstairs, sitting or standing [45].

The entire process is composed by a pipeline of four phases described in the following:

- **Pre-training process:** The HAR is trained to detect the correct user action on a big public dataset (HMOG dataset).
- **Data collection:** The inertial data of the user are collected by the smartphone application in unobtrusive way.
- **Enrollment phase:** The HAR is fine-tuned with the data of the user.
- **Feature extraction:** A set of representative user walking features are extracted from the last layers of the HAR and stored in a database.
- **Similarity check:** The walking features collected when the passenger is in the transportation hub and the set of walking representative features stored are compared with a similarity metric (e.g., Similarity with Dynamic Time Warping (DTW)).

2.4.3. Data Format Requirement

The data expected to train the machine learning classifier and provide the similarity check would take as input the temporal sequence of the logs of the inertial data of the accelerometer (Acc) and gyroscope (Gyr sensors along the three axes x, y, z.

Timestamp	Acc x_axis	Acc y_axis	Acc z_axis	Gyr x_axis	Gyr y_axis	Gyr z_axis
252207918580802	-4.332779	13.361191	-0.7188721	-0.85321045	0.29722595	0.8901825
252207968934806	-0.31944275	13.318359	-0.23202515	-0.8751373	0.015472412	0.16223145
252208019288809	1.566452	9.515274	-0.01777649	-0.72016907	0.38848877	-0.28401184

2.4.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-GA-DM-01	MUST	The gait analysis system provides interaction through the edge device (smartphone) and the cloud service for the walking path analysis.	E-CORRIDOR Ope-01
E-CORRIDOR-GA-DM-02	MUST	The cloud service, after the authentication process must delete the subject information.	E-CORRIDOR-DS-10
E-CORRIDOR-GA-DM-03	MUST	The gait analysis system allows to control the access to every transportation hub through the walking path user verification.	E-CORRIDOR-DS-11

2.4.5. Application to Pilots

Pilot	Airport-Train pilot
Reference to Use cases or User stories	• AT-US-02: Frictionless Multimodal Journey • AT-US-03: Distributed and Combined Context Analysis in Sensor Network • AT-US-05: End to End Safe-Contact/Contactless Journey • AT-US-07: Document-free Secure Multimodal Travel Credential • AT-UC-13: Privacy-aware Behavioral Identification
Brief description of the Use cases or User stories	The above use cases aim to make easy the travel of the user reducing the multiple user interactions with the transportation check-in systems ensuring minimum disruption in the context of authentication process, identification of the user documents and identify.
Match of the proposed approach/technology with the USs/UCs	The gait analysis approach is a continuous and unobtrusive authentication mechanism that exploit inertial sensors embedded in the smartphone user. Such unobtrusiveness can be exploited to reduce the user interaction during the authentication process in every transportation hub.

2.4.6. Potential Synergies

Synergies with other components - Work package and Task	T8.1
Title/brief description of the task	Privacy aware seamless multimodal authentication
Description of the potential synergy with risks and opportunities	In the context of the seamless multimodal authentication, the gait analysis can contribute to reach the multimodal authentication mechanism thanks to the background analysis of the inertial data during the walking path.
Dependencies on other components	None

2.5. Face recognition- passenger authentication [E-CORRIDOR-IAI-FR]

In recent decades, interest in face recognition theories and algorithms has grown quickly. The following are just a few examples of concrete applications that use face recognition algorithms and that have gained attraction among industries: video surveillance, criminal identification, building access control, and autonomous vehicles. Different techniques, including local, holistic, and hybrid approaches, which provide a face image description using only a few face image features or the whole facial features, are developed. Since face recognition is a process of identifying or verifying the person's identity using his/her face, it is usually used in many

applications with consumers outside of smartphones services, such as airport check-ins, sports stadiums, and concerts. The basic steps of the face recognition process are as follows: i) face detection, ii) face capture, and iii) face match. The first step is essential in detecting and locating human faces in images and videos. In the second step, analog information (a face) is transformed into a set of digital information (data) based on the person's facial features. The third step is essential in verifying if two faces belong to the same person. Recognition of passengers' faces with a high accuracy will increase the accuracy of their authentication by expanding the set of their contextual information. Developing a face recognition system with a high degree of robustness and discrimination poses several challenges, such as head orientation, lighting conditions, and facial expression. Moreover, face recognition systems usually need high processing time, high memory consumption and are relatively complex.

2.5.1. State of the Art

Face recognition is a challenging topic in several application domains, including video surveillance, criminal identification, building access control, and autonomous vehicles. In addition, many car companies are experimenting with face recognition approaches. One of their objectives is to use face recognition approaches to replace a face with a key for starting a car. Another objective is to change radio stations and seat preferences based on the driver. Face recognition can also increase drivers' safety by recognizing and alerting drivers if they are not focusing on the road.

The existing face recognition approaches usually use different sensors, including RGB, depth, Electroencephalography (EEG), thermal, and wearable inertial sensors. There are three groups of sensors that may improve the reliability and the accuracy of a face recognition system: i) non-visual sensors, ii) detailed-face sensors, iii) target-focused sensors. Non-visual sensors, such as audio, depth, and EEG sensors, provide extra information, e.g., illumination variation and position shift situation, in addition to the visual dimension and improve the reliability of the recognition. Detailed-face sensors, such as eye-trackers, detects a small dynamic change of face, which may help distinguish background noise and face images. Target-focused sensors, such as infrared thermal sensors, can make the filtering of useless visual contents easier and may help resistance illumination variation [46].

Figure 6 shows the face recognition structure, which includes the mentioned three basic steps: i) face detection, ii) face capture, and iii) face match.

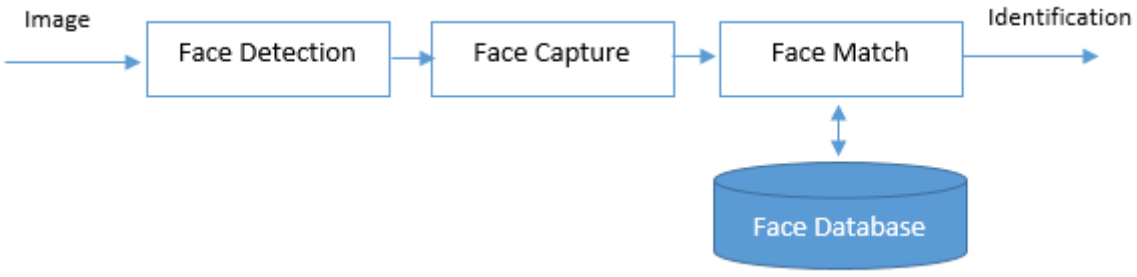

Figure 6. Face recognition structure

In the literature, there are many techniques for face detection, such as Viola–Jones detector [47], [48], histogram of oriented gradient (HOG) [49] [50], and principal component analysis (PCA) [51], [52]. Moreover, the face detection step can be exploited for video and image

classification, object detection [53], region-of-interest detection [54]. Face capture techniques, such as HOG [55], Eigenface [56], independent component analysis (ICA), linear discriminant analysis (LDA) [51] [57], Scale-Invariant Feature Transform (SIFT) [58], gabor filter, local phase quantization (LPQ) [59], Haar wavelets, Fourier transforms [60], and local binary pattern (LBP) [61] [62] techniques are widely used to extract the face features. Correlation filters (CFs) [63] [64] [65], convolutional neural network (CNN) [66], and also k-nearest neighbor (K-NN) [67] are known techniques for face matching.

The existing face recognition approaches can be classified into three main categories: i) local, ii) holistic (subspace), and iii) hybrid. In the first category, the recognition is performed based on the certain facial features, not considering the whole face. Approaches related to the second category use the whole face as input data and then project it into a small subspace or in correlation plane. In hybrid approaches, local approaches and holistic ones are combined to deal with their limitations while exploiting their advantages. The most commonly used local approaches for face recognition are LBP [68], Histogram of oriented gradients (HOG) [69], correlation filters (CFs) [49], SIFT [58], Speeded-up robust features (SURF) [53], Binary robust independent elementary features (BRIEF) [54], Fast retina keypoint (FREAK) [70]. Although, these approaches provide robust recognition under different illumination conditions and facial expressions, they are sensitive to noise, and invariant to rotations [46].On the other hand, the most commonly used holistic approaches for face recognition are Eigenface [56] and principal component analysis (PCA) [71], Fisherface and LDA [72], ICA, Gabor filters [73], Discrete wavelet transform (DWT) [74], Discrete Cosine Transform (DCT) [75], Gabor-KLDA [76], Kernel PCA (KPCA) [52], and Kernel Linear Discriminant Analysis (KDA) [77]. These approaches allow a better reduction in dimensions and an improvement in the recognition rate, however, they are not invariant to translations and rotations compared with local techniques. Hybrid approaches combine local and holistic approaches to offer better performance for face recognition systems. Some examples of hybrid face recognition approaches are as follows:

- Gabor wavelet and linear discriminant analysis (GW-LDA) [78]
- Over-complete LBP (OCLBP), LDA, and within class covariance normalization (WCCN) [79]
- Advanced correlation filters and Walsh LBP (WLBP) [80]
- Multi-sub-region-based correlation filter bank (MS-CFB) [81]
- CNNs and stacked auto-encoder (SAE) techniques [82]
- PCA and ANFIS [83]
- DCT and PCA [84]
- PCA, SIFT, and iterative closest point (ICP) [85]
  - PCA, local Gabor binary pattern histogram sequence (LGBPHS), and GABOR wavelets [86]
  - PCA and Fisher linear discriminant (FLD) [87] [88]
  - SPCA–KNN [89]
  - Convolution operations, LSTM recurrent units, and ELM classifier [90]

2.5.2. Proposed Approach/Technology

In the E-CORRIDOR framework, a machine learning based mechanism will be defined to implement a face recognition tool usable in every kiosk located in the train station and airport. The tool exploits biometrics to map facial features from photographs coming from a set of cameras. For instance, the proposed approach captures the locations and outlines of each user's eyes, nose, mouth, and chin to model the user's face. It then compares the collected biometrics information with a database of known faces. As it is mentioned, face recognition can assist verification of personal identity. The proposed face recognition tool is described in Figure 7:

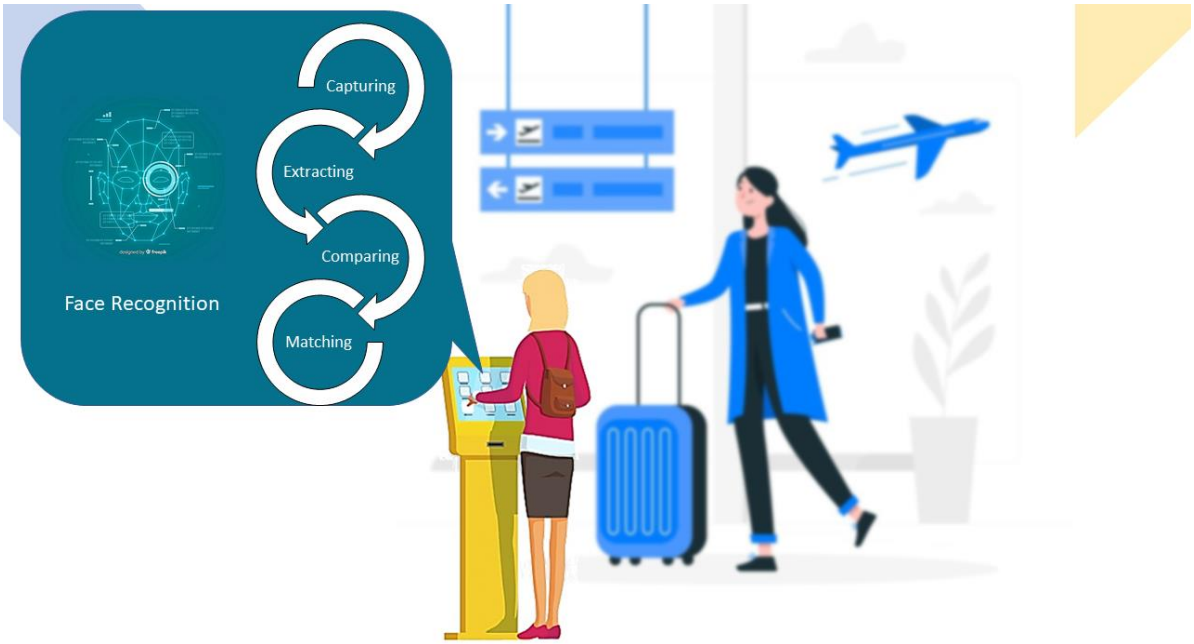

Figure 7. The proposed face recognition tool

The first component is capturing the images using cameras. The second one is extracting the unique facial data from the captured images. The third component is comparing; in this component, the facial data is compared with the database. The last component is matching which decides whether the sample (new image) matches with any images in the face database or not.

The core of this tool is a Deep Neural Network (DNN) called residual neural network, which was introduced for image recognition [91]. After the process of face detection using Histogram of Oriented Gradients (HOG), the residual neural network is exploited for face recognition.

In the context of the E-CORRIDOR project, for the sake of the privacy of users, there is no persistent face database of passengers and the processes take place at the edge of the system. To this end, we consider collecting pictures of the passengers' face at the first checkpoint, e.g., where the passenger enters train station or airport. Then, the model is trained using the collected pictures. We plan to use a hybrid approach based on the DNN technique and HOG. In the checkpoints after the initial collection of face information, the model is verified against the pictures collected in the current checkpoint and eventually updated. The passenger's facial information can be destroyed once the passenger leaves the airport to preserve privacy. It is worth mentioning that our proposed model learns the features useful for face recognition incrementally. The entire process is composed of the following steps:

1- Collect face database at the first checkpoint
2- Train face recognition model including face detection, face capture, and face match at the second checkpoint
3- Retrain face recognition model at the following checkpoints

2.5.3. Data Format Requirement

In input the component will take color images with depth information (in RGB-D format). The produced output will be a Boolean value expressing if the detected face matches the one detected at the previous touch point.

2.5.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-FR-01	MUST	The accuracy and effectiveness of passengers' face recognition is dependent on the DSA specified by each passenger, passengers' activity, and on contextual/environmental properties.	• E-CORRIDOR-DS-05 • E-CORRIDOR-DS-06 • E-CORRIDOR-DS-07 • E-CORRIDOR-DS-17 • E-CORRIDOR-DS-23 • E-CORRIDOR-DS-24
E-CORRIDOR-IAI-FR-02	MUST	Face recognition can be performed at the edge.	• E-CORRIDOR Ope-02
E-CORRIDOR-IAI- FR-03	MUST	IP connected camera and Light Detection and Range camera are used for face recognition to identify passengers	• E-CORRIDOR-Tst-AT-02 • E-CORRIDOR-Tst-AT-03
E-CORRIDOR-IAI-FR-04	SHOULD	The face recognition of each passenger should be used to enhance the seamless authentication.	• E-CORRIDOR-Use-02
E-CORRIDOR-IAI-FR-05	MUST	The inferred passenger face information is transmitted and stored (only for the needed time) in a privacy-aware and secure manner.	• E-CORRIDOR-DS-10

2.5.5. Application to Pilots

Pilot	Airport-Train pilot
Reference to Use cases or User stories	• AT-UC-01: PRM Passenger Assistance and Authorization • AT-UC-02: Passenger and Baggage Contextual Identification • AT-UC-03: Contactless Passenger Authentication and Authorization • AT-UC-04: Privacy-preserving Passenger Monitoring • AT-UC-06: Single Sign-On (SSO) Authentication • AT-UC-12 Passenger Flow Overview and Prediction • AT-UC-13 Privacy-aware Behavioral Identification • AT-US-01: Passenger Management and Operations • AT-US-03: Distributed and Combined Context Analysis in Sensor Network • AT-US-05: End to End Safe-Contact/Contactless Journey • AT-US-07: Document-free Secure Multimodal Travel Credential
Brief description of the Use cases or User stories	The above use cases and user stories refer to situations in which the passenger is moving within the premise of the airport.
Match of the proposed approach/technology with the USs/UCs	The data are used to recognize the face of passengers inside the airport. The data will also be used to perform flow assessments in the airport to support end-to-end safe-contact/contactless journeys.

2.5.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.1 • T8.3
Title/brief description of the task	The above tasks refer to privacy aware interest-based service sharing, seamless multimodal authentication, and the passenger's contextual authentication.
Description of the potential synergy with risks and opportunities	The data gathered from the passenger have to be shared within the framework respecting the privacy policies set by the passenger itself. The same information can also be used to expand the contextual behavioral authentication of the passenger.
Dependencies on other components	The sensor fusion and the context reasoning can provide a strong multi-biometric and seamless multimodal authentication.

2.6. Activity recognition- passenger authentication [E-CORRIDOR-IAI-AR]

Human activity recognition can be defined as the process of determining and naming activities using data collected from wearable, environmental, or vision sensors. In detail, human activity refers to the movements of one or more parts of the person's body. The technology of Human activity recognition (HAR) has motivated the development of various context-aware applications in emerging domains, e.g., the Internet of Things (IoT), Ambient Assisted Living (AAL), and healthcare. HAR analyzes data acquired from different types of sensing devices, including vision sensors or/and embedded sensors. In the E-CORRIDOR project, the objective of using the HAR component is to obtain contextual information to enhance the authentication obtained by other components, such as face recognition and gait analysis.

2.6.1. State of the Art

The existing HAR approaches can be classified into two main categories: i) sensor-based HAR and ii) vision-based HAR. The sensor-based HAR approaches concentrate on investigating raw data extracted from wearable sensors and environmental sensors. In contrast, the vision-based HAR approaches analyze images or videos obtained from optical sensors [92]. Since not all the passengers will be willing to wear sensors (or share data from their wearable devices), the HAR component of this project is vision-based. Vision-based HAR approaches rely on visual sensing technologies, such as CCTV and camera, to record human activities [93]. These approaches depend on the quality of images, including image resolution, lighting environments, and illumination changes.

In [94], 3D and depth data are used for the recognition of human activities. 3D skeleton-based human representation and activity recognition approaches are studied in several works [95] [96] [97]. Indeed, spatiotemporal human representation based on 3D visual perception data is a rapidly growing research area. In general, spatiotemporal human representations can be classified into two main categories depending on whether they use RGB-D information or 3D skeleton data. Due to skeleton-based human representations' robustness to variations of viewpoint, human body scale, motion speed, and real-time, online performance, these approaches have attracted an increasing attention [95]. 3D skeleton-based representations allow modeling the relationship of human joints and encode the whole body configuration [95].

For obtaining 3D skeleton data, there are several commercial devices, such as motion capture systems, time-of-flight sensors, and structured-light cameras. Figure 8 shows several 3D skeletal kinematic human body models provided by the different devices; e.g., the OpenNI library tracks 15 joints; Kinect v1 SDK tracks 20 joints; and the Kinect v2 SDK tracks 25 joints.

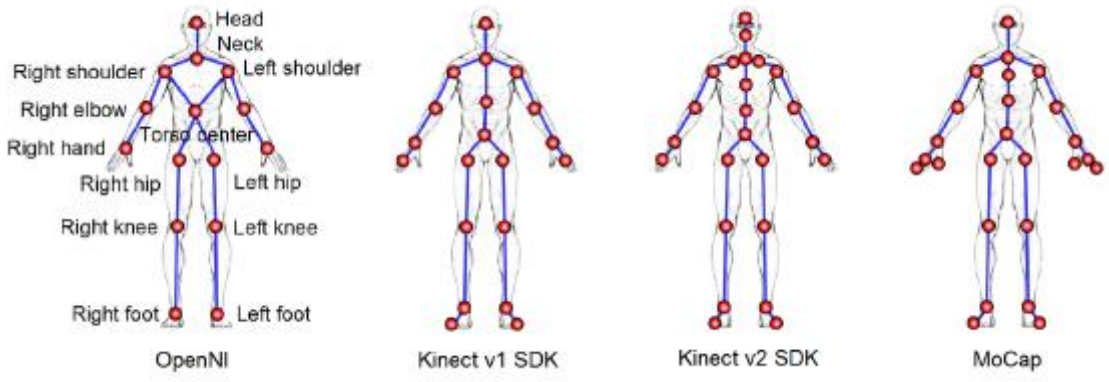

Figure 8. Several skeletal human body models obtained from different devices [95].

Recently, deep learning methods are extensively used for skeleton-based human representations. Similar to using deep learning methods for feature extraction from images where raw pixels are typically used as input, skeleton-based human representations constructed by deep learning methods generally rely on the raw joint position information. In [98], an end-to-end hierarchical Recurrent Neural Network (RNN) for the skeleton-based representation construction is proposed. In this study, the raw positions of human joints are directly used as the input to the RNN. In [99], raw 3D joint coordinates are used as input for an RNN with Long Short-Term Memory (LSTM) to learn human representations automatically.

2.6.2. Proposed Approach/Technology

In the E-CORRIDOR framework, the activity recognition component is based on a deep learning method to model human activities using 3D action sequences. Human actions can be considered as time series of configurations of skeletal data of users. The latter can be modeled using the 3D locations of major joints of the users' bodies. In other words, each sample is represented as a sequence of these configurations. RNNs and LSTMs have been used to learn sequential data in different applications. One of the main limitations of traditional RNNs is their inability to keep the long-term representation of the sequences, making them unable to find relations among long ranges of inputs. To deal with this limitation, LSTM was introduced; LSTMs keep a long-term memory inside each RNN unit and learn when the information stored inside its internal memory cell should be remembered or forgotten.

In the proposed activity recognition, we plan to use a part-aware LSTM human action learning model. Body joints move in groups in human actions, while each group can be mapped to a major part of the body. Interactions between body parts or with other objects can be used to interpret actions. In the proposed part-aware LSTM human action learning model, we don't keep a long-term memory of the entire body's motion in the cell, but we split it into part-based cells. In other words, we keep the context of each body part independently; there are individual input, forget, and modulation gates for each part's cell; however, the output gate is shared among the body parts. The output of the model can be seen as a combination of context information of independent body parts. In the proposed model, the body joints are grouped into five parts: torso, right hand, left hand, right leg, and left leg. At each frame, we concatenate the 3D coordinates of the joints within each part and use them as the part's input representation.

One of the main characteristics of our activity recognition model is needing limited training parameters, which avoid the overfitting problem. This can be explained by the fact that the traditional LSTM has full connections between all the memory cells and input features using an input modulation gate. Moreover, the memory cell is used to represent the long-term dynamics of the entire skeleton over time, resulting in a large number of training parameters prone to overfitting. In our activity recognition model, unnecessary links are dropped to deal with this problem since the entire body's dynamics, represented in the memory cell, are divided into the dynamics of body parts. The proposed model learns the common temporal patterns of the body parts independently and then combines them to recognize activities.

Figure 9 shows the architecture of the activity recognition model. One can observe that each body part has its individual input, forget, and modulation gates while the output gate is shared between body parts.

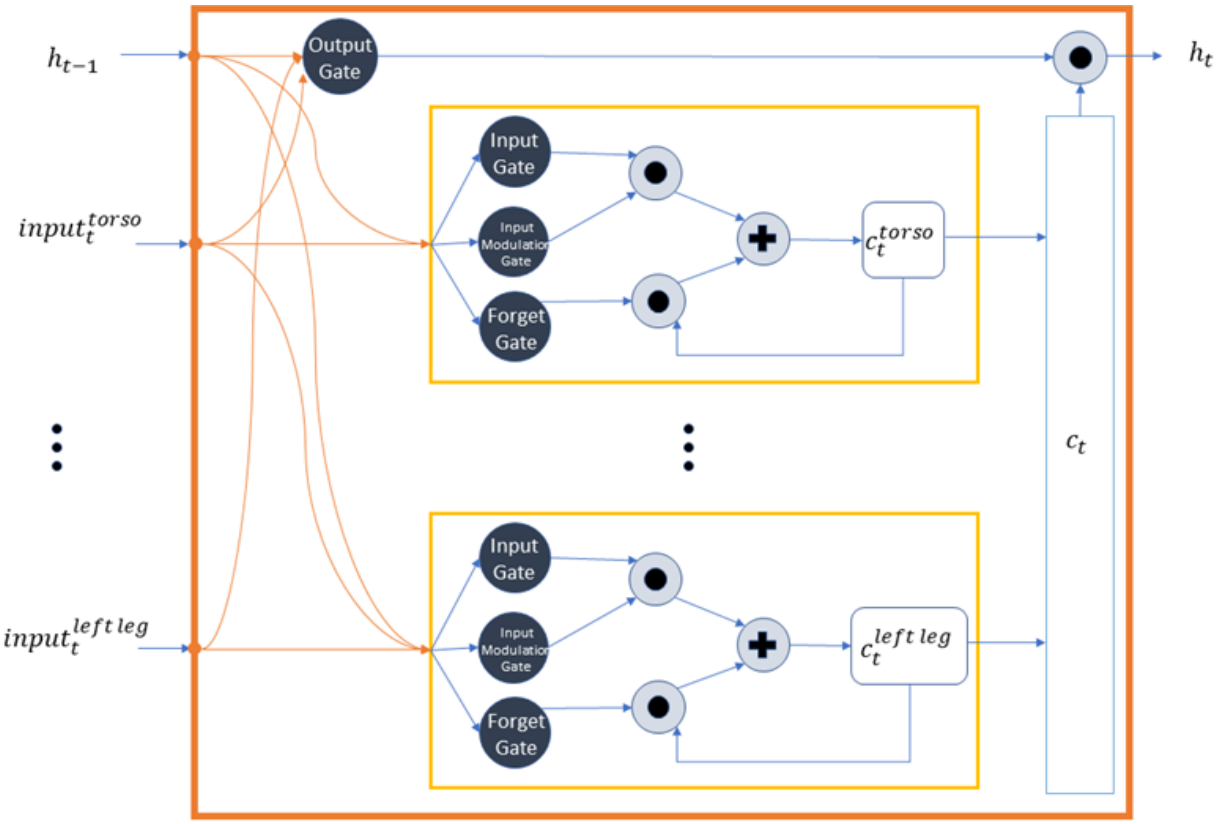

Figure 9. Architecture of the activity recognition model

2.6.3. Data Format Requirement

For each passenger, sequences of RGB videos with depth information are collected. Moreover, 3D skeletal data and infrared (IR) videos can be exploited. The resolutions of RGB videos are expected to have a resolution of at least 1920x1080, whereas depth maps and IR videos of 512x424. Three dimensional skeletal data are expected to contain the 3D coordinates of 25 body joints at each frame. In output the component will produce the label of the corresponding activity.

2.6.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-AR-01	MUST	The accuracy and effectiveness of passengers' activity recognition is dependent on the DSA specified by each passenger, and on contextual/environmental properties.	• E-CORRIDOR-DS-05 • E-CORRIDOR-DS-06 • E-CORRIDOR-DS-07 • E-CORRIDOR-DS-17 • E-CORRIDOR-DS-23 • E-CORRIDOR-DS-24
E-CORRIDOR-IAI-AR-02	MUST	Activity recognition can be performed at the edge.	• E-CORRIDOR Ope-02
E-CORRIDOR-IAI- AR-03	MUST	IP connected camera and Light Detection and Range camera are used for activity recognition to identify passengers	• E-CORRIDOR-Tst-AT-02 • E-CORRIDOR-Tst-AT-03
E-CORRIDOR-IAI-AR-04	SHOULD	The activity recognition of each passenger should be used to enhance the seamless authentication.	• E-CORRIDOR-Use-02
E-CORRIDOR-IAI-AR-05	MUST	The inferred passenger activity information is transmitted and stored (only for the needed time) in a privacy-aware and secure manner.	• E-CORRIDOR-DS-10

2.6.5. Application to Pilots

Pilot	Airport-Train pilot
Reference to Use cases or User stories	• AT-UC-01: PRM Passenger Assistance and Authorization • AT-UC-02: Passenger and Baggage Contextual Identification • AT-UC-03: Contactless Passenger Authentication and Authorization • AT-UC-04: Privacy-preserving Passenger Monitoring • AT-UC-06: Single Sign-On (SSO) Authentication • AT-UC-12 Passenger Flow Overview and Prediction • AT-UC-13 Privacy-aware Behavioral Identification • AT-US-01: Passenger Management and Operations • AT-US-03: Distributed and Combined Context Analysis in Sensor Network • AT-US-05: End to End Safe-Contact/Contactless Journey • AT-US-07: Document-free Secure Multimodal Travel Credential
Brief description of the Use cases or User stories	The above use cases and user stories refer to situations in which the passenger is moving within the premise of the airport.
Match of the proposed approach/technology with the USs/UCs	The output of the activity recognition analysis is used to characterize the activity of passengers inside the airport and the train station. The data will also be used to perform flow assessments in the airport to support end-to-end safe-contact/contactless journeys.

2.6.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.1 • T8.3
Title/brief description of the task	The above tasks refer to privacy aware interest-based service sharing, seamless multimodal authentication, and the passenger's contextual authentication.
Description of the potential synergy with risks and opportunities	The data gathered from the passenger have to be shared within the framework respecting the privacy policies set by the passenger itself. The same information can also be used to expand the contextual behavioral authentication of the passenger.
Dependencies on other components	The sensor fusion and the context reasoning can provide a strong multi-biometric and seamless multimodal authentication.

3. Privacy Preserving Itinerary Planning – Task 7.2

Task 7.2 - Privacy Preserving Itinerary Planning - is dedicated to the design, implementation and maturation of analytics to infer or predict the best multi-modal travel itineraries for end-users (e.g., passengers, mobility service users, drivers). The analytics will consider users' interests and preferences, the CO2 footprint of the possible itineraries, price, time and number of connections. The analytics designed in this task should be able to use anonymized data, to not hinder the user's privacy. Also, they need to be self-adaptive, recomputing the itinerary at runtime, according to possible context change or critical situations on the initial itinerary.

Itinerary planning will be an important data analytics functionality for people to securely access the multi-modal transport service within the E-CORRIDOR framework and transport operators to evaluate the performance of their provided transport service (e.g., service coverage). In this section, we focus on defining the requirements for a CO2-aware trip planning data analytics tool that will be integrated into IAI, considering the E-CORRIDOR scenarios and pilot use cases.

3.1. CO2-aware Trip Planning [E-CORRIDOR-IAI-MMIP]

The CO2-aware Trip Planning data analytics tool will explore approaches for integrating promising mobility solutions (e.g., electric vehicle car sharing and on-demand bus service) with public transit to enhance individuals' mobility. End users can plan their trips and gain access to innovative shared or on-demand mobility concepts and solutions, through the trip planning tool provided by E-CORRIDOR. Moreover, personal trip preferences such as transport modes and carbon footprint will be considered when planning trips, to enhance the overall user experience and contribute in realizing societal goals (e.g., European Green Deal).

In this subsection, we will first review the state-of-the-art trip planning technologies and tools, and then propose E-CORRIDOR's approach to designing this data analytics tool and applying it to E-CORRIDOR pilots.

3.1.1. State of the Art

In this subsection, we have conducted a survey to compare the features of several commercial off-the-shelf online trip planners and identify the features that we aim to deliver within E-CORRIDOR. Specifically, we extract the evaluation metrics mainly from the use cases defined by the project, their popularity, and the support for open source. Hence, six online trip planning tools have been chosen, and the comparison result is presented in the following table. Besides, it should be noted that this comparison matrix is not exhaustive, considering the number of trip planning tools we have surveyed [100], and just keeps the most representative ones.

Table 1. A brief comparison of several popular trip planning tools.

Tool Name	Open Source	Supported Transport Modes	Map Data	GTFS support	Real-time Traffic	CO2 estimation
Google Map/ Transit [101]	No	car, bicycle, walk, transit	Proprietary	Yes	Yes	No
TripGo [102]	No	car, bicycle, walk, transit, taxi, car-sharing and more.	OSM	Yes	Yes (use the live traffic from Google or TomTom)	Yes
OpenTripPlanner (OTP) [103]	Yes	car, bicycle, walk, transit	OSM	Yes	Partially (GTFS-Realtime)	No
Open Source Routing Machine (OSRM) [104]	Yes	car, bicycle, walk	OSM (and NED)	No	No	No
OpenRouteService (ORS) [105]	Yes	car, bicycle, walk, transit (bus)	OSM	No	No	No
GraphHopper [106]	Yes	car, bicycle, walk, transit (bus), truck, scooter	OSM	Yes	No	No

These trip planners can be classified as open-source and closed source ones, when considering their support for open source. The former category is represented by Google Map/ Transit and TripGo, while the latter is represented by OpenTripPlanner (OTP). These commercial closed-source trip planners have a stable performance and wide applications, but are hard for testing new features and integrating with other solutions. However, both can support multiple transport modes and real-time traffic and service information, and TripGo even supports CO2 estimation. Thus, there is a trend for the trip planner to become smarter, greener, and more real-time.

For open-source trip planners, all of them use the OpenStreetMap (OSM) [107] data for routing and support several common transport modes. It should be noted that OTP supports more transit modes (e.g., bus, tram, subway) and has implemented lots of experimental features through its Sandbox [108]. Regarding the support for General Transit Feed Specification (GTFS) [109], which is a data specification that defines public transportation schedules and associated geographic information, OTP supports both GTFS and GTFS-Realtime [110], enabling it to consider both the static and real-time service information for route calculations. Last, none of these trip planners officially support carbon footprint estimation or prediction, due to the focus of their development work.

All in all, by conducting this survey, we have seen some great features from these trip planners, such as the support for open-source communities, being real-time and user-friendly, and the willingness of raising awareness of carbon footprint from transport domains.

3.1.2. Proposed Approach/Technology

The CO2-aware Trip Planning data analytics tool will be mainly based on the OpenTripPlanner 2 (OTP2) due to its remarkable characteristics of supporting open source, multi-modal transport, and GTFS (including GTFS-Realtime and GTFS-Flex). The tool will allow end-users

to define an origin and destination within a specific pilot region, and receive personalized multi-modal itineraries which both include flexible transport services provided by E-CORRIDOR partners and consider users' interests and preferences (such as the CO2 footprint of the possible itineraries, price, time and number of connections).

The following figure shows the architecture of OTP2, which is adapted from Figure 4 in [111] by referencing and analyzing the source codes of OTP2 on GitHub. To be specific, the core of the OpenTripPlanner architecture is a routing engine written in Java that finds efficient paths through multi-modal transportation networks built from OpenStreetMap (OSM) and GTFS data. The Routing API is a RESTful web service that responds to trip planning requests with returned itineraries in a JSON or XML representation, while the Graph Visualizer provides a JavaScript-based front end to show the map and trip planning options. Additionally, Graph Visualizer also needs to invoke Graph Builder to obtain the map data. However, the OTP instance can also work without a graphical user interface by calling the Routing API directly. Thus, RESTful APIs to use the core functionalities of the component will be provided through the IAI. The front end will be customize only if required by the considered use case.

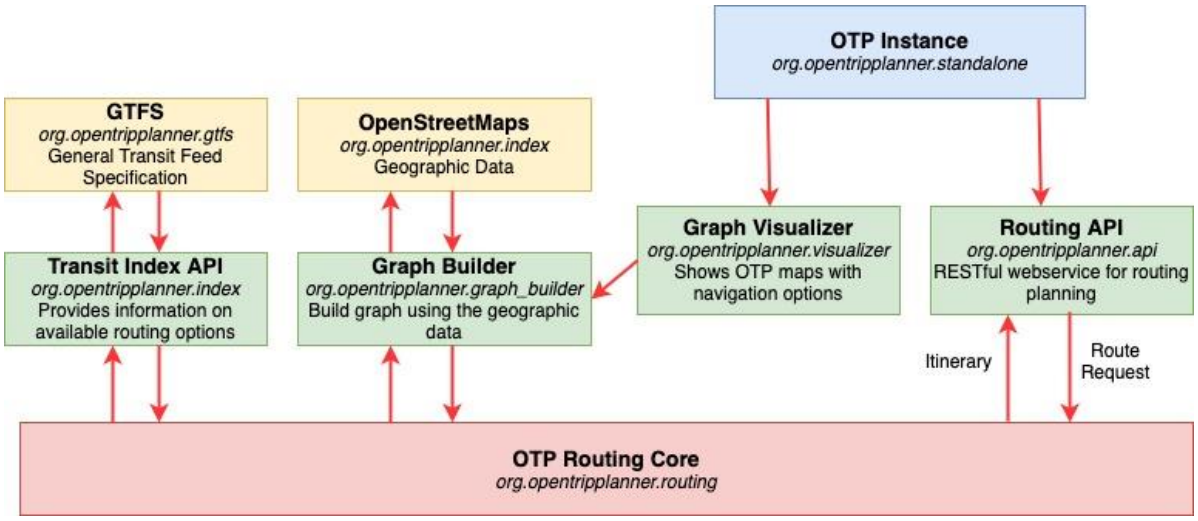

Figure 10. OTP2 architecture, adapted from [111].

The OTP Routing Core involves two separate services for its proper functionality, which are OTP Transit Index API and Graph Builder. The Transit Index API is a RESTful web service that reads GTFS feeds and feeds the information into the routing core. GTFS contains public transportation schedules and associated geographic data and is a vital data source for supporting multi-modal trip planning. Furthermore, Graph Builder makes a graph for representing road networks, out of various open-source geographical map data such as OSM. Map data is essential for the proper operation of the OTP Routing Core.

To deliver a versatile CO2-aware trip planning tool, the core research and development (R&D) work will be conducted around the following aspects:

- Modification of OTP Routing Core to support new mobility modes and services (mainly car-sharing and on-demand bus service originated from WP3) and new routing options (least CO2 emission).
- Enhancing the Transit Index API to include CO2 profiles of various transportation modes to empower further CO2 estimation.
- Improve the available privacy preservation features.

For the third aspect, the trip planning tool can decide whether to log all the incoming (trip planning) requests for later analysis. This is an optional feature since some transit operators and agencies may use the logs to identify existing or unmet transportation demand. A log will contain the following eight fields: 1) date and time the request was received; 2) IP address of the user; 3) arrive or depart search; 4) arrival or departure time; 5) all transport modes selected; 6) origin latitude and longitude 7) destination latitude and longitude 8) travel duration in seconds and the number of transit vehicles used in that itinerary returned to the user. It should be noted that most of these fields are essential for trip planning requests and should be supplied by users when using the tool. Nonetheless, they are not linked with the identities of users since the trip planning tool does not require any login.

In future development efforts of our component, we will also consider using anonymized user information (provided by prosumers shared through the ISI) to support the automation of trip planning. An E-CORRIDOR prosumer can define in DSAs which data can be shared with the trip planning tool and how the trip planning tool can use the data needed for planning a trip (e.g., current location and timestamp). However, the trip planning tool still needs the cooperation from prosumers since the data fields such as arriving or departing and destination locations should be indicated by users. If the data shared in ISI contains more private data such as the preferences of transport modes, we will also consider using the FHE tool provided by the E-CORRIDOR framework to avoid accessing user data directly. All in all, the proposed trip planning tool will consider the privacy issues in its lifecycle and utilise the privacy preservation tools of the E-CORRIDOR framework to resolve privacy concerns.

With the introduction of new mobility modes and service, we adapt the search logic of the routing engine to accommodate the needs originated from E-CORRIDOR pilots and user cases. However, we will remain adopting the same routing algorithms Generalized cost A* [112] and RAPTOP [113]) due to their reasonable performance.

3.1.3. Expected Data Format

In this subsection, the public datasets or data feeds to be used by the itinerary planning data analytics will be listed and explained. The itinerary planning data analytics tool needs to build a transit network and a road network during the bootstrap stage, and this lays the foundation for further routing work. The expected data here refers to the transit and map data needed by the itinerary planning tool to build a transportation network (also called “graph”).

GTFS and GTFS-Realtime (for bus service providers such as Pildo in the S2C pilot of the E-CORRIDOR project)

The General Transit Feed Specification (GTFS) is a data specification that allows public transit agencies to publish their transit data in a format that can be consumed by a wide variety of software applications. [109]. Besides, GTFS include a static component, GTFS-Static [114], which is mandatory and contains schedule, fare, and geographic transit information and a real-time component, GTFS-Realtime [110], which is optional and contains arrival predictions, vehicle positions and service advisories.

Public transit agencies need to provide their transit data in the GTFS format to allow itinerary planning tools to consume that data in an interoperable way and utilize accurate and even real-time transit information for routing calculations. Besides, GTFS should be prepared by following the relevant specifications and fully validated, before publishing them on the web for public access.

For the E-CORRIDOR use cases (mainly S2C-UC-03), it should be noted that to support on-demand business service, *Shapes* contained in a file *shapes.txt*, which describe the path that a vehicle travels along a route alignment, should be provided. Relevant requirements can be referenced in [115].

General Bikeshare Feed Specification (GBFS) [116]

The General Bikeshare Feed Specification (GBFS) is the open data standard for shared mobility. Similar to GTFS, GBFS makes real-time shared-mobility data feeds in a uniform format and publicly available online. It should be noted that even though the "Bikeshare" may imply GBFS was originally proposed for bike-sharing, GBFS does support other shared mobility service such as car-sharing and scooters.

Car-sharing service providers such as (Clem' in WP3) should prepare their GBFSs in order to make the transit information (such as car-sharing stations, the number and types of available cars) available for trip planning tools. The current release for GBFS is v2.2, and this version should be respected in E-CORRIDOR's implementation.

3.1.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-MMIP-01	SHOULD	The multi-modal trip planning tool should be able to pull data (such as public transit feeds) from external sources, with specified polling intervals.	• E-CORRIDOR-DS-20 • E-CORRIDOR-Tst-Int-S2C-02
E-CORRIDOR-IAI-MMIP-02	MUST	Data used to automate trip planning needs to be obfuscated or anonymized to enhance the privacy preservation. It should be also deleted after a certain amount of time.	• E-CORRIDOR-DM-01 • E-CORRIDOR-DM-02 • E-CORRIDOR-Sec-RC-01 • E-CORRIDOR-DS-10

3.1.5. Application to Pilots

Pilot	Car-sharing pilot
Reference to Use cases or User stories	• S2C-US-05: Trip planning and Carbon footprint. • S2C-UC-03: Trip planning and carbon footprint analysis
Brief description of the Use cases or User stories	The above use case/user story refers to a scenario where travellers plan to calculate optimized routes for their multimodal trips according to their criteria and check relevant trip information, and calculate and track carbon footprint info of their trip both before and after trips.
Match of the proposed approach/technology with the USs/UCs	To support the multi-modal transport service within the E-CORRIDOR framework, a privacy-preserving and versatile trip planner will be a traveler-oriented service within the E-CORRIDOR IAI Analytics Toolbox. OTP 2 can meet most of the demands derived from the US/UC and is open to support more advanced trip planning features with further R&D work.

3.1.6. Potential Synergies

Synergies with other components - Work package and Task	• Task 7.4 Carbon foot print analytics
Title/brief description of the task	T7.4 aims at designing analytics for inferring by approximation, with a limited knowledge of all involved elements, the actual CO2 footprint in multi-modal transport system. This task will provide analytics which can estimate the CO2 footprint according to information acquired in real time, such as adjusting travel time, driving style, fuel quality, etc. These analytics will be performed in a privacy preserving manner, using anonymized or generalized data, or differential privacy.
Description of the potential synergy with risks and opportunities	The analysis of carbon footprint relies significantly on the trip information. The transport vehicle used and travelling distance are the two most important factors for CO2 estimation. By designing more accurate CO2 estimation algorithms within T7.4 and integrating these algorithms with our trip planning tool, we can better present the carbon footprint information of potential trips to travellers and allow them to choose more carbon-free transport solutions.
Dependencies on other components	Carbon foot print analytics in Task 7.4

4. Privacy Preserving (Security) Analytics – Task 7.3

The main objective of this task is to define a generic OpenAPI platform which allow to easily integrate privacy preserving analytics to be applied on shared cybersecurity information. The task will also design and implement secure analytics exploiting Homomorphic Encryption, extending the CEA (partner of the E-CORRIDOR project) crypto-computing compiler, and analytics which exploit other privacy preserving techniques such as anonymization, generalization and differential privacy. The analytics developed will be based on both statistical analysis for vulnerability and attack correlation, exploiting cascade analysis and machine learning for attack pattern recognition, malware and text (email) analysis, network-based attack detection. This task will put a particular attention in developing analytics which have a good trade-off between accuracy and ensured privacy, also aiming at minimizing performance overhead.
In this section, we propose two approaches for privacy preserving analytics:

- OpenAPI for Fully Homomorphic Encryption
- Private Secure Routine

4.1. OpenAPI for Fully Homomorphic Encryption [E-CORRIDOR-IAI-FHEC]

The Fully Homomorphic Encryption (FHE) Analytics is a part of E-CORRIDOR Analytics Toolbox. It is based on Cingulata which is a source to source compiler developed by CEA[1], and BigPi platform used for applying homomorphic cryptographic techniques which on top of allowing the scrambling of data in order to protect its confidentiality also provides the necessary mathematical building blocks for performing privacy-preserving calculations, by the execution of general algorithms directly on encrypted data. It is described in Figure 11.

[1] https://github.com/CEA-LIST/Cingulata

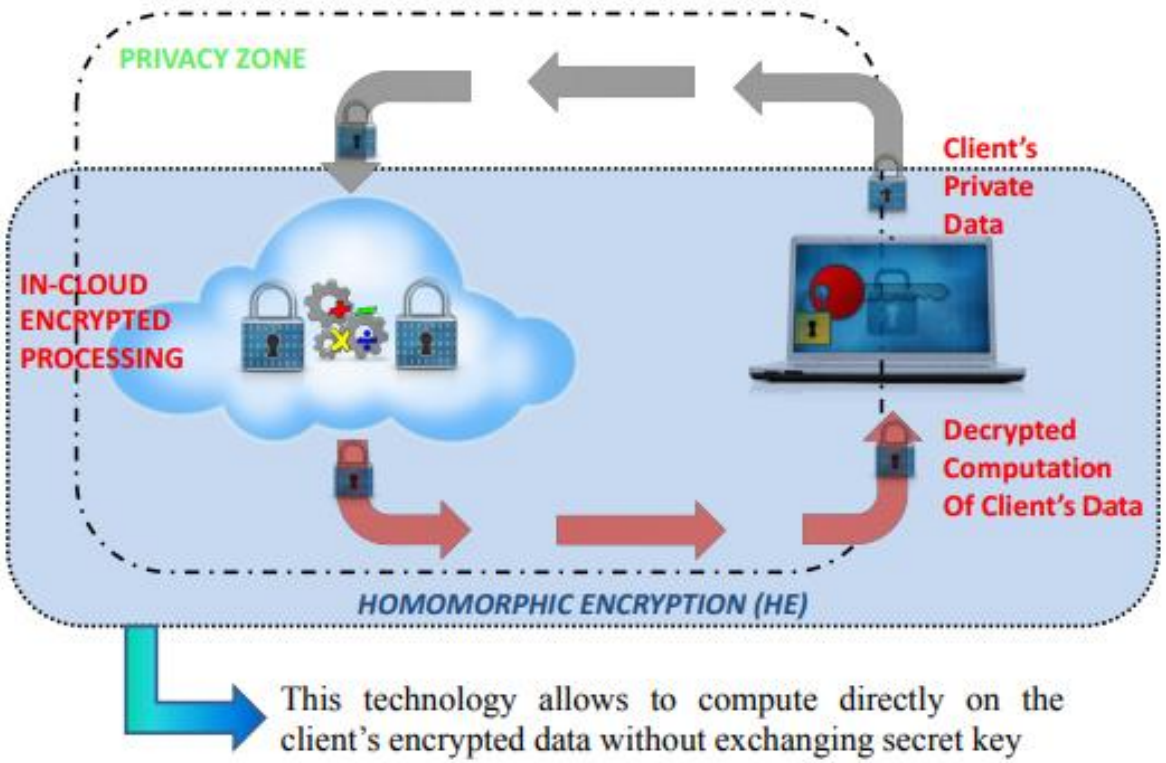

Figure 11: High level view of BigPi platform

The BigPi platform is based on Cingulata toolchain. This toolchain architecture is pictured in Figure 12. It consists mainly of the following components:

- A compiler infrastructure for high-level cryptocomputing-ready programming, taking C++ code as input.
- Boolean circuit optimization, parallel code generation and «cryptoexecution» runtime environment.
- Optimized prototypes of the most efficient homomorphic encryption systems known so far.

In Figure 12, Cingulata first compiles an application written in C++ language into a Boolean circuit (blue area). Whereupon, Cingulata optimizes the circuit to improve performance (red area). To save bandwidth, transcryption can be performed with a homomorphic-friendly standard cryptosystem (green area). In the end, the circuit is evaluated over ciphertexts using a homomorphic scheme (red area).

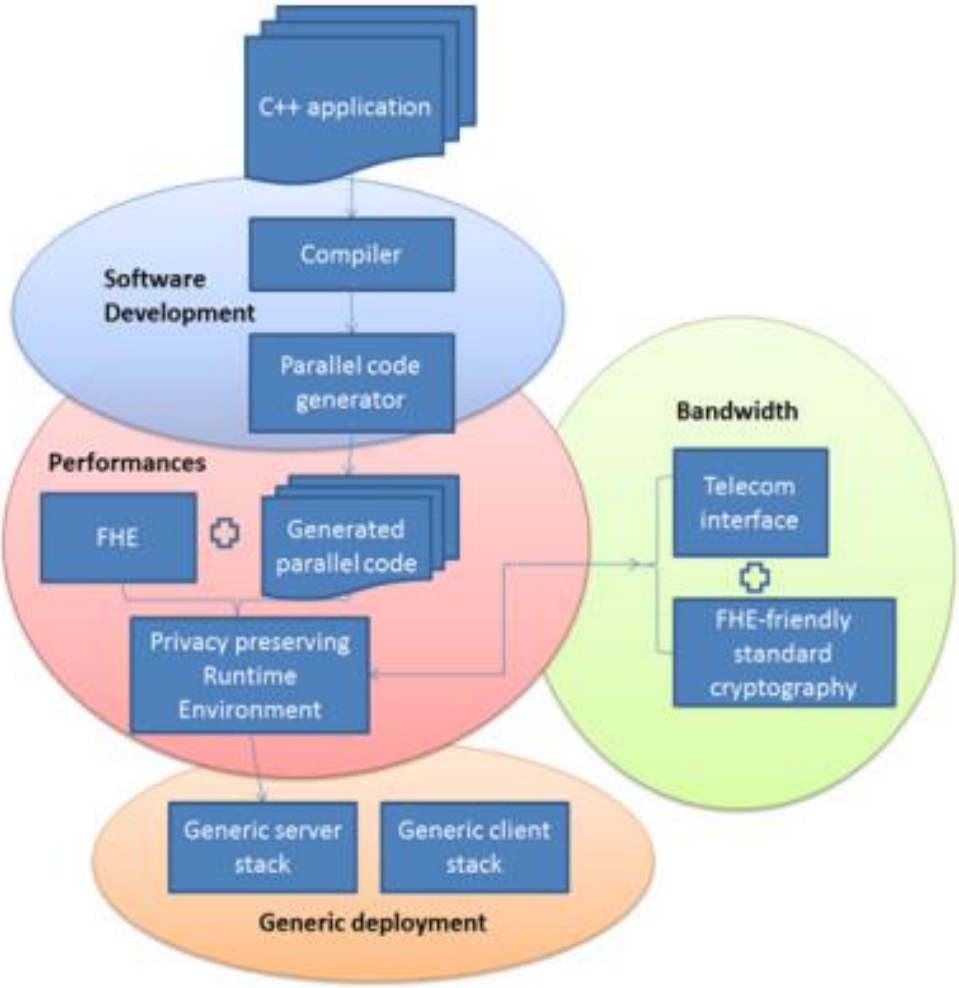

Figure 12: Cingulata Architecture

4.1.1. State of the Art

Homomorphic encryption (HE) is a recent cryptographic method allowing performing computation directly on encrypted data, without the need of decrypting it. As such, the encryption schemes possessing homomorphic properties can be very useful to construct privacy preserving protocols, in which the confidential data remains secured not only during the exchange and the storage but also for the processing. In a context of data outsourcing and of cloud computing, the homomorphic encryption is a mechanism that helps to protect data from intrusions from the cloud provider itself. The service provider (cloud) processes the received data homomorphically and sends the encrypted result to the end user, owner of the homomorphic secret key.

In real world cloud applications using FHE encryption, one or several entities interact with the cloud and, to preserve the privacy of each user, their data are sent encrypted over the cloud. The service provider processes the received data homomorphically and sends the encrypted result to an end user (owning the FHE parameters and, hence its secret key). The latter one decrypts the result using its own decryption key. Here, the service provider can compute almost any functions over the encrypted data and acts transparently with respect to each entity using only public information and homomorphic encrypted data.

In order to address the practicality issues, we dispose nowadays of several tools and methods to bring to reality homomorphic-based cloud applications. There are several FHE schemes quite efficient (each one with its advantages and disadvantages) as well as several open-source libraries implementing it (e.g., SEAL[2], PALISADE[3] or TFHE[4]). Moreover, it exists a theoretical framework (Chimera) allowing to switch between these different cryptosystems in order to choose the most appropriate for various parts of the computation in the homomorphic domain. The CEA team has worked on the design, development and maintenance of the open-source Cingulata compiler environment (https://github.com/CEA-LIST/Cingulata), the first operational tool of this kind. The integration of TFHE (standing for Fast Fully Homomorphic Encryption over the Torus and belonging to the 3rd generation of FHE schemes) into Cingulata compilation chain was realized in June 2019. As such, Cingulata offers the possibility to execute Boolean circuits, either with BFV cryptosystem (and thus the execution is dependent of the multiplicative depth) or with TFHE (only *13ms* to perform a gate evaluation) techniques, in the E-CORRIDOR platform and provides an added – value of enhanced privacy – protecting framework. Developing and adopting Cloud – first deployment strategy, the secure sharing approaches based on homomorphic encryption help ensuring data confidentiality while allowing secure processing.

4.1.2. Proposed Approach/Technology

Homomorphic encryption (HE) is an encryption method which allows to perform computation on encrypted data without decrypting it. Such schemes are known to be very useful to construct privacy preserving protocols even in its classical version. As example, homomorphic encryption has been used as a key-tool in the popularization of electronic-based voting scheme. Another application of homomorphic encryption is Private Information Retrieval, which is a communication efficient interactive protocol which allows a user to retrieve an item in a database without revealing which item he is looking for. This paradigm has found a number of applications in numerous contexts: private searching, keyword search, private storage, anonymous authentication, etc. Another very popular scenario which makes the benefits of homomorphic encryption is cloud computing: a user relies on some computing resources from a cloud provider to perform expensive computation on sensitive data. These scenarios have in common that Fully HE (FHE) encryption is used as a method which allows the scrambling of data in order to protect their confidentiality via the execution of algorithm on encrypted data. In real world applications using FHE encryption, one or several entities interact with the cloud. To preserve privacy of each user, the data are sent encrypted over the cloud. The service provider processes the received data homomorphically and sends the encrypted result to an end user (owning the FHE parameters and, hence its secret key). The latter one decrypts the result using its own decryption key. Here, the service provider can compute almost any functions over the encrypted data and acts transparently with respect to each entity using only public information and encrypted data.

[2] https://github.com/microsoft/SEAL

[3] https://github.com/gchq/Palisade

[4] https://github.com/tfhe/tfhe

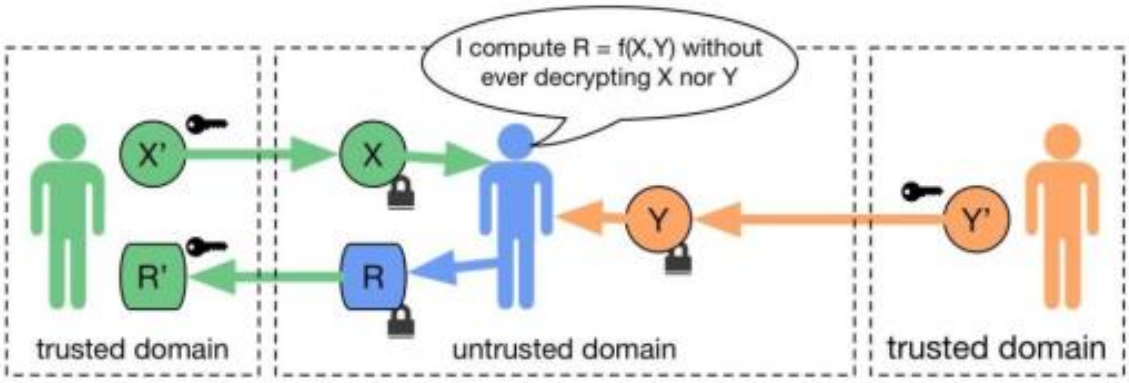

Figure 13: Homomorphic encryption allows computation on encrypted data without decrypting it in untrust environments.

The underlying mathematical objects used to conceive fully homomorphic encryption schemes are Euclidean lattices. The security of almost all known FHE construction relies on the problem of finding short vector or basis in a high dimensional lattice. Gentry's solution relies on ideal lattices over algebraic number fields. In 2012, Brakerski, Gentry and Vaikuntanathan [117] improved this scheme without using bootstrapping; they proposed a generalized construction secure under the popular Learning With Errors assumption and its ring variant. Then Brakerski [118] proposed a new scale invariant scheme that does not require modulus switching. In 2012, Fan and Vercauteren (FV) [119] proposed a ring variant scheme and improved its efficiency. The so-called BGV and FV cryptosystems which are already implemented in version Cingulata 1.0 [120]. The 3rd generation of FHE with fast bootstrapping techniques called TFHE - Fast Fully Homomorphic Encryption over the Torus based on [121] [122] is released in the version Cingulata 2.0 [120] since June 2019.

Our Cingulata open source version offers a compiler chain with high-level language development targeting HE execution based on manipulating Boolean circuits. That is a directed graph *G = (V, A)* which vertices are either inputs, outputs or operators *(XOR, AND)* and which arcs corresponds to data transfers. The following constraints are imposed to a compiler targeting HE execution [123] [124]:

- No *if* conditions (unless regularized by conditional assignment).
- No data dependent loop termination (it needs upper bounds).
- Array dereferencing/assignment in *O (n) (vs O (1)).*
- Algorithms always realize (at least) their worst-case complexity!

In terms of technology design, this compilation chain is composed of three layers: a front-end, a middle-end and a back-end. The front-end transforms code written in C++ into its Boolean circuit representation. The middle-end layer optimizes the Boolean circuit produced by the front-end. The back-end homomorphically executes the Boolean circuit over encrypted data. Two HE libraries are supported by our Cingulata compiler: (i) an in-house implementation of [119] and (ii) the publicly available TFHE library.

A simple "hello world" example written using Cingulata is:

```
CiInt a{CiInt::u8};      // create an unsigned 8-bit variable
CiInt b{CiInt::u8v(42)}; // use helper function to create
```

```
                          // an unsigned 8-bit
CiInt c{-1, 16, false};  // or manually specify value,
                          // size and signedness
a.read("a");             // read variable a and b
b.read("b");

c = a + b;
```

Using the FV cryptosystem this program is homomorphically executed in less than 5 seconds and using TFHE in less than 1 second, whereas by applying BFV in Z^n with matrix multiplication and bootstrapping (an option available in the SEAL library), the program is completed after 0.5 second.

In the E-CORRIDOR project, we use the *Pattern Searching* functionalities. The algorithm is quite simple thanks to the FHE performance and allowed operations. For reducing computation times, a pre-processing should be executed on the client side in order to extract the principal text to analyze, referred as *T*. Using a list of sensible patterns which are already encrypted in homomorphic format and stored in a database, for each encrypted pattern, the component seek and check whether those pattern are present in *T*. From our last experiments, the check of 5000 encrypted sensible patterns is completed in less than 1 second.

In the context of the S2C pilot, checking, in a privacy preserving manner, the validity of a driving license in a collaborative database of multiple car sharing actors from different countries constitutes an interesting use case. The performance of the *PaternSearch* scheme have been tested in an implementation exploiting 40 threads and applying the SEAL 3.5.1 library. Analysis and decryption algorithms have been executed with 3 different data sets respectively constituted by 4.000, 8.000 and 16.000 encrypted driving licenses. The table below shows the average time (in second) of 3 executions of each algorithm. The time for encrypting a vector is negligible and therefore it is not considered in this simple benchmark.

# of driving licenses	Evaluation time (second)	Decryption time (second)	Total (in second)
4.000	0.477	0.061	0.46
8.000	3.591	0.116	3.7
16.000	39.257	0.256	39.5

4.1.3. Data Format Requirement

To use and facilitate encrypting data in FHE format, the data input in plaintext is provided in the format below:

```
Prefix Data_Value
Index Data 1; Data Value in plaintext
…
Index Data n; Data Value in plaintext
```

Where the prefix is an index constituted by any alphanumeric sequence.

In case of the driving licenses discussed before the input format was constituted by:

```
1; FR-13-090413302170-PTH
2; FR-75-190475302143-LKM
3; FR-92-490492302132-CDR
```

4.1.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil D5.1 Requirement(s)
E-CORRIDOR-IAI-FHEC-01	MUST	The Mapper functionality grants Prosumers that the translation of data sharing constraints is compliant and consistent from the high level to the low level specification.	• E-CORRIDOR-DA-03 • E-CORRIDOR-DA-05
E-CORRIDOR-IAI-FHEC-02	MUST	Data used to identify drivers may require to be transformed into a common data format to work with the analytics.	• E-CORRIDOR-DA-02
E-CORRIDOR-IAI-FHEC-03	MUST	The In-Vehicle Infotainment (IVI) or Electronic Control Units (ECUs) may be used for collecting GPS and connection behaviour data.	• E-CORRIDOR-Tst-S2C-01 • E-CORRIDOR-Tst-S2C-02

4.1.5. Application to Pilots

Pilot	AT, S2C, ISAC pilot,
Reference to Use cases or User stories	• AT-US-06: De-silo and Co-optimize Operations Data • AT-US-04: Advanced Security Analytic Services • ISAC-US-03: ISAC-MMT cyber-threat information analysis • ISAC-US-04: ISAC-MMT cyber-threat notification • ISAC-US-07: sharing automotive cyber-threat information analysis • ISAC-US-06: Automotive cyber-threat information analysis • S2C-US-01 : eWallet • S2C-US-06 Cybersecurity notifications: communicate about threats • S2C-US-07 Secure sensitive data that would be shared from end to end
Brief description of the Use cases or User stories	This user stories refer to the access of sensitive data and the execution of cyber-security analysis.
Match of the proposed approach/technology with the USs/UCs	The adoption of this analytics will allow the E-corridor framework to identify connection IP in black list or white list, or text analysis.

4.1.6. Potential Synergies

Synergies with other components - Work package and Task	• WP8 – T8.3
Title/brief description of the task	The above tasks refer to Privacy Aware Interest-Based Service Sharing
Description of the potential synergy with risks and opportunities	The FHE API could constitute a building block for preserving the privacy while providing services based on the user interests.
Dependencies by other components	None

4.2. Secure Multiparty-computation for Routine based authentication - Private Secure Routine [E-CORRIDOR-IAI-MPCSR]

Vehicles circulating on roads generate huge amount of data about both the driver and the vehicle itself. Such data can be used for different purposes, e.g., data generated may indicate the type of driving style or used to identify drivers. However, when processed, these data may reveal sensitive information. So, they should be processed with respect to drivers' privacy.

We propose Private Secure Routine (PSR) as a paradigm with two main objectives: i) identify drivers depending on their habits/routine and ii) keep private drivers' data. We implemented PSR exploiting the Secure Multi-Party Computation (MPC) technique against a honest-but-curious attacker model.

4.2.1. State of the Art

In literature there are several solutions based on machine learning and neural network (NN) techniques for driver identification. Among them, we report the ones that we consider more relevant with respect to the Private Secure Routine.

Micale et al. [125] propose Secure Routine (SR), a paradigm that identifies drivers according to their routines and way of drive. The authors use Random Forest algorithm and define a procedure to select the features that better represent each driver. SR was tested on datasets [126], with a precision of 99,8% and a recall of 98,5% and [127], with a precision of 99,6% and a recall of 98,1%.

Martinelli et al. compared different Decision Trees algorithms on dataset Θ using all features on the research [128] and using only the six best features in [129]. Authors obtained up to 99,2% of precision and recall using J48.

Uvarov et al. [130] highlight the issue of car manufacturers that use non standard IDs of sensors' data of the CAN messages. It is not always possible to obtain the databases with the IDs information of each vehicle. Hence, authors verified how accurate can be driver identification models using only public sensors' data available with every OBD-II dongle. In the experiments, they use the dataset Θ and removed every feature not publicly available. Authors' best result is

79% of accuracy using Random Forest in multi-driver identification, i.e., identify who is actually driving the car, whereas on the owner identification problem authors obtained 99% of accuracy.

Feng et al. [131] predict human mobility by using Federated Learning technique. Vehicles work together to create a model with the help of a server. Each vehicle customizes the model using a "personal adaptor" to better predict personal mobility patterns.

Costantino et al. [15] propose a driver reputation characterization calculated in a privacy preserving way by using secure Multi-Party Computation. They collect vehicles' sensor data to calculate the Reputation score. The authors describe some example of ITS services that can be customized according to the reputation of the driver. The reputation score is calculated without machine learning.

4.2.2. Proposed Approach/Technology

The Private Secure Routine (PSR) is a paradigm to identify drivers belonging to the same vehicle in a privacy-preserving manner. Private Secure Routine is built on top of the Secure Routine (SR) paradigm [125]. The advantages of PSR are twofold:

- PSR is able to distinguish among several drivers depending on their routine. While SR is able to identify only one driver for a target vehicle, referred to as the owner of the vehicle, PSR is able to identify more than one authorized driver for a target vehicle.
- PSR guarantees that information about drivers and vehicles are exchanged in a privacy-preserving way by exploiting the secure multi-party computation technique.

The PSR paradigm takes as input all the pieces of information about drivers and vehicles circulating in the infrastructure and generates models of each driver in each vehicle. This is made by combining Federated Learning [132] and Secure Multi-Party Computation (MPC) [133], [134] protocol, as cryptography technique.

Federated Learning (FL) allows multiple peers to generate a common model, (e.g., a neural network), without sharing their data in order to overcome critical NN training issues, e.g, data privacy. Secure Multi-Party Computation is a cryptography technique that involves n parties, where each party i holds the input x_i, and all participants want to compute a function f (x_1, x_2, ..., x_n) maintaining private each party input. The function f in a secret sharing scheme is randomly split into n secrets, named shares, in such a way that certain subsets of shares can be used to reconstruct the secret and others reveal nothing about it.

The Private Secure Routine paradigm is implemented by using PySyft framework, which is a Python library that implements the secure Multi-Party Computation (MPC) technique for private training of Neural Networks [135]. The framework maintains both parameters of the model and the dataset private. Note that the PySyft implementation of MPC is secure against the honest-but-curious adversaries [135] but cannot guarantee security against active attackers. Some parties could exchange their shares and potentially reconstruct the original values.

To identify drivers in a vehicle, the Private Secure Routine paradigm creates a model able to identify each driver of a target vehicle that circulates on the PSR infrastructure. Model generation depends on different situations that can occur and involve both drivers and vehicles.

We assume that all communications among vehicles and infrastructure happen through secure channels. This will overcome possible fully malicious attacks. Also, we use an asymmetric cryptography protocol [136].

4.2.3. Data Format Requirement

There is not a required format of data. It depends on the in-vehicle and environmental sensors. Hence, data can be CAN messages collected from the OBD-II port or interacting directly with the in-vehicle network. E.g., from the OBD-II port information such as GPS coordinates, RPM and fuel consumptions can be retrieved and exploited by the component. The output is a Boolean flag stating if the driver has been identified as an authorized driver or not.

4.2.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-MPCSR-01	MUST	Stakeholders may require running analytics expressing conditions to preserve confidentiality over the shared data.	• E-CORRIDOR-DS-09
E-CORRIDOR-IAI-MPCSR-02	MUST	Data used to identify drivers may require to be obfuscated, anonymized or other privacy-preserving technologies must be adopted.	• E-CORRIDOR-DM-01 • E-CORRIDOR-Sec-RC-01
E-CORRIDOR - IAI-MPCSR-03	SHOULD	Data used to identify drivers may require to be transformed into a common data format to work with the analytics.	• E-CORRIDOR-DA-02
E-CORRIDOR - IAI-SR-02	MUST	Driver DNA analytics can be run at the edge.	• E-CORRIDOR Ope-02
E-CORRIDOR - IAI-SR-03	SHOULD	The In-Vehicle Infotainment (IVI) or Electronic Control Units (ECUs) may be used for collecting GPS and driving behaviour data.	• E-CORRIDOR-Tst-S2C-01 • E-CORRIDOR-Tst-S2C-02

4.2.5. Application to Pilots

Pilot	S2C, ISAC
Reference to Use cases or User stories	• S2C-US-9: Driving behavior recognition • ISAC-US-01: Public cyber-threat information collection
Brief description of the Use cases or User stories	The above use cases refer to the analysis of the driving behavior. The same input data can be collected for identifying potential cyber-threat information
Match of the proposed approach/technology with the USs/UCs	The adoption of this analytics will allow the E-corridor framework to identify drivers according to their driving styles.

4.2.6. Potential Synergies

Synergies with other components - Work package and Task	• T8.2
Title/brief description of the task	The above tasks refer to continuous behavioral authentication.
Description of the potential synergy with risks and opportunities	By using PSR it is possible to identify all drivers that are authorized to drive a target vehicle. In order to authorize a driver it is important to identify and consequently authenticate her. This can be done by enhancing the current version of PSR with some authentication method.
Dependencies on other components	Secure routine for driver identification (driver DNA)

5. Carbon Footprint Analytics – Task 7.4

Task 7.4 Carbon Footprint Analytics aims at designing analytics for inferring by approximation, with limited knowledge of all involved elements, the actual CO2 footprint in the multi-modal transport system. The accurate calculation of the carbon footprint of a process (e.g., a trip) requires a large set of precise information, which is difficult to collect. Thus, this task will provide analytics that can estimate the CO2 footprint according to information acquired in real time, such as the travel distance, driving style, fuel quality, etc. These analytics will be performed in a privacy-preserving manner, using anonymized or generalized data, or differential privacy.

5.1. CO2 analytics [E-CORRIDOR-IAI-CFA]

The Carbon Footprint Analytics will provide a tool and accompanying algorithms to estimate the CO2 footprint of trips, according to limited information acquired in real-time, such as the travel distance, driving style, fuel quality, etc. The carbon footprint information of trips will then be returned to end-users (such as passenger and transport operators), to allow them to know the carbon emissions of using or running multi-modal transport service.

In this section, we will first review the state-of-the-art CO2 analytics technologies and tools, and then propose E-CORRIDOR's approach to designing this data analytic tool and applying it to E-CORRIDOR pilots.

5.1.1. State of the Art

Transport represents almost a quarter of Europe's greenhouse gas (GHG) emissions and is the main cause of air pollution in cities. Within the transport sector, road transport is by far the biggest emitter accounting for more than 70% of all GHG emissions from transport in 2014 [137].

According to the research led by John Mulrow [138], most people recognize the significant influence of transportation activities on their carbon footprint calculation, and they are generally more curious about the carbon footprint generated by transportation (than home energy, food, water, and others). Designing carbon footprint calculators for transportation contributes to better estimation of the carbon emission originated from transportation-related activities and realization of the EU's targets for reducing greenhouse gas (GHG). This research also provides a detailed comparison among 31 popular carbon footprint calculators provided by government organizations, non-profit organization and private companies. However, all these calculators adopt survey-based methods to acquire relevant inputs from users, and none of them supports automatic CO2 calculation or inference, when given the details of a trip. Also, few of them support multi-modal transport.

Another trend that we have noticed is that more trip planning tools are supporting carbon footprint estimation, such as TripGo [102] by SKEDGO and the Green Driving Tool [139] by Joint Research Centre (JRC). CO2 analytics seems a natural match with trip planners, since providing CO2 estimation before a trip will affect the travel choice, and trip planners can provide important data such as travel distance, transport modes and geographical information. Furthermore, one important feature we have learned from the Green Driving Tool is the concept of Citizen Science, where people can provide fuel consumption and routes to help researchers to build more accurate models to analyze CO2 emissions. The idea of Citizen Science is somehow similar to the information sharing concept adopted by E-CORRIDOR, and by using

advanced data analytics techniques, great insights and more comprehensive models can be obtained.

5.1.2. Proposed Approach\Technology

Within this task, a carbon footprint calculator will be developed to allow users to estimate the carbon footprints of their multi-modal trips. Users need to provide the travel distance, transport mode and vehicle type to get the CO2 estimation. It should be noted that as the research and development work evolves, more complicated inputs (such as driving style analyzed in E-CORRIDOR-IAI-SR see Section 2.1) may be needed to get a more accurate calculation.

The following figure shows the architecture of the CO2 calculator for multi-modal trips. First, the CO2 Calculation API is a RESTful web service that responds to CO2 calculation requests with returned CO2 information in a JSON representation. Second, CO2 Calculation Core will contain the logic and algorithms for CO2 calculation. Most of the research and development effort will be spent on optimizing the CO2 calculation algorithms that suit the E-CORRIDOR use cases. Last, a Carbon Profile Database will be linked with the Core to check the carbon profiles of different transport modes and vehicle types. A database can avoid hard-coded CO2 values and provide high expansibility and maintainability.

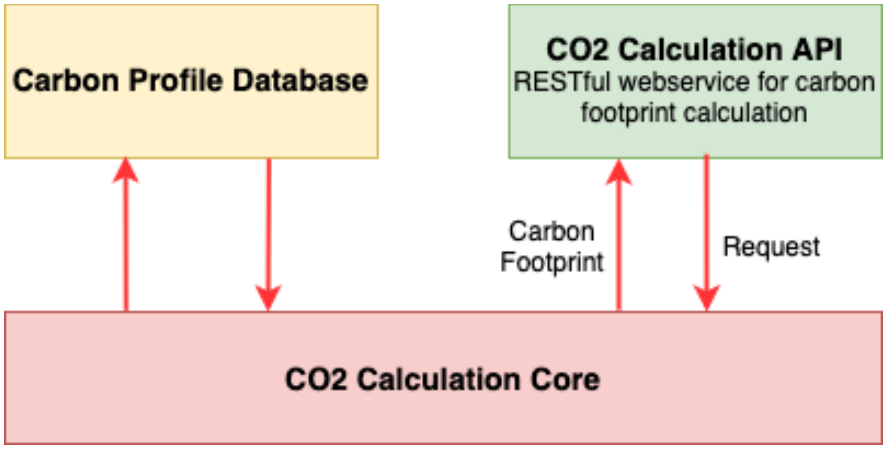

Figure 14. CO2 analytics architecture.

5.1.3. Expected Data Format

In order to estimate the carbon footprint of trips, an end-user needs to provide some essential information such as distance, transport mode, and vehicle type. There is no standardized format for use due to the various features and functionalities of different carbon footprint analytics tools. Thus, the data needed for CO2 calculation should be contained in a proprietary format to be discussed with the smart cities and car-sharing (S2C) pilot partners, Clem' and Pildo. Also, the data format will evolve with the development of the Carbon Footprint Analytics and be expanded to allow different levels of accuracy and data inputs. Initially, the CO2 Analytics tool needs **distance, transport mode,** and **vehicle type** for a basic CO2 calculation.

5.1.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-CFA-001	SHOULD	The carbon footprint analytics tool should be able to pull data (such as carbon profiles of different vehicles) from external sources, with specified polling intervals.	• E-CORRIDOR-DS-20 • E-CORRIDOR-Tst-Int-S2C-02

5.1.5. Application to Pilots

Pilot	Car-Sharing pilot
Reference to Use cases or User stories	• S2C-US-05: Trip planning and carbon footprint. • S2C-UC-03: Trip planning and carbon footprint analysis
Brief description of the Use cases or User stories	The above use case/user story refers to a scenario where travelers plan to calculate optimized routes for their multimodal trips according to their criteria and check relevant trip information and calculate and track carbon footprint info of their trip both before and after the trips.
Match of the proposed approach/technology with the USs/UCs	The proposed technology can estimate the carbon footprint of multi-modal trips, when given the distance travelled, transport types, and vehicle types. The CO2 calculator can be integrated with trip planners to automatically calculate the carbon footprint or run in a standalone mode by exposing APIs to end-users.

5.1.6. Potential Synergies

Synergies with other components - Work package and Task	• Task 7.2 Privacy preserving itinerary planning • Task 7.1 Data analytics for driver identification
Title/brief description of the task	Task 7.2 is dedicated at the design, implementation and maturation of analytics to infer or predict the best multi-modal travel itineraries for end-users. Task 7.1 gathers and processes a variety of data produced from cars (OBD readings, GPS, etc.), transport entities, users, infrastructure's detectors, sensors, and even social media using the novel analytics E-CORRIDOR platform. This task leverages advance artificial intelligence algorithms to allow driver identification, authentication and possibly driving style.
Description of the potential synergy with risks and opportunities	Task 7.2 E-CORRIDOR-IAI-MMIP: The analysis of carbon footprint relies greatly on the trip information. By designing more accurate CO2 estimation algorithms within T7.4 and integrating these algorithms with our trip planning tool, we can better present the carbon footprint information of potential trips to travelers and allow them to choose more carbon-free transport solutions. Task 7.1 E-CORRIDOR-IAI-SR: Driving style (if available) could be considered as another factor affecting the CO2 calculating results, and Task 7.4 Carbon Footprint Analytics will conduct research to include these kinds of factors into the carbon footprint calculation and provide more accurate results.
Dependencies on other components	The itinerary planning tool in Task 7.2

6. Intrusion Detection Technologies – Task 7.5

This task will provide security analytics utilizing machine learning for anomaly-based intrusion detection. In particular, the following important capabilities will be provided for multi-modal transport applications: behavior conformance tracking; security compliance tracking, and prediction of critical situations. Behavior conformance tracking is the capability to detect deviations of observed events from expected events with respect to the multi-modal transport application model and the current state. Security compliance tracking is the capability to apply a security model at runtime in order to identify violations of security requirements. Prediction of critical situations is the capability to predict violations of security requirements in the near future. More precisely, if a state transition leads to a critical state in a security monitor within the behavior prediction scope then a so-called predictive alert will be raised. We provide a multi-dimensional behavioral detection engine model that includes operational, system, and network data to detect advanced correlated attacks. The developed behavioral detection engine relates states of different transport systems so that attacks can be identified with higher confidence level. The proposed approach adapts relations in order to capture the nature of attacks. Moreover, Advance Persistent Threats can be easily detected with this learning approach.

In the following, we describe the requirements and architecture for the components developed in this task, in particular with respect to intrusion prevention and detection systems.

6.1. Automotive Intrusion Detection [E-CORRIDOR-IAI-CANIDS]

In this subsection, we will describe the E-CORRIDOR requirements and architecture for the specific components to be used for the intrusion detection with data from vehicular networks.

In the past, cars could only be tampered if someone had direct physical access. With connected cars, however, we are now in an era where the technology exists for attackers to remotely target millions of vehicles simultaneously. Protecting connected vehicles will require holistic approaches to design, implement, and respond when the unexpected does happen. The attack surface of a modern car consists mostly of networked components and physical access to the internal communication bus systems. These channels could be used to inject and transmit adversarial messages, e.g., over the controller area network (CAN), or Automotive Ethernet, or other external interfaces, to an electronic control unit (ECU). In the future generation of connected cars and multi-modal transport systems in general, new attack vectors arise due to the increasing automation of vehicular functions. A potential attack scenario would be IP network attack to get unauthorized access to safety-related advanced driver-assistance systems (ADAS) controls of a vehicle. Those attacks can produce direct and controllable functional loss or impact on functional safety, which makes it one of the most prominent threats in cyber security.

6.1.1. State of the Art

Despite the trend to use Automotive Ethernet in recent in-vehicle architectures, CAN bus is still in use and can be utilized to attack modern vehicles [140]. Research on vehicular IDS has also focused almost exclusively on CAN traffic. The CAN intrusion detection methods can be sorted in four categories:

1) Detecting specification violations.
2) Detecting ECU impersonating attacks
3) Detecting packet insertions.

4) Detecting sequence context anomalies.

With respect to category 1 (specification violations), [141] describe a set of network-based detection sensors, which allow the recognition of anomalies occurring inside the vehicular network. These are characterized by two categories, namely specification-based sensors and semantic-based sensors (see Table 2).

Table 2 List of network-based detection sensors

Sensor	**Description**
Formality	Correct message size, header and field size, field delimiters, checksum, etc.
Location	Message is allowed with respect to dedicated bus system.
Range	Compliance of payload in terms of data range.
Frequency	Timing behavior of messages is approved.
Correlation	Correlation of messages on different bus systems adheres to specification.
Protocol	Specification Correct order, start-time, etc. of internal challenge-response protocols.
Plausibility	Content of message payload is plausible, no infeasible correlation with previous values.
Consistency	Data from redundant sources is consistent.

[142] proposes specific checks, e.g. for formality, protocol and data range. [143] describes among others a specific frequency sensor. [144] describes a language-based intrusion detection approach which could be seen as an extension of the protocol sensor by adding the specification of the state-machines of the participants to the protocol checks. [145] describes specific semantic technologies that could be used for plausibility and consistency sensors.

These methods cannot detect attacks that act within the specified ranges but they have the advantage of avoiding false positives, therefore industrial products also often use this kind of rule-based IDS (e.g. https://www.escrypt.com/en/products/cycurids).

With respect to category 2 (ECU impersonation), the standardized automotive open system architecture (AUTOSAR) specified a module for secure onboard communication (SecOC) to check the authenticity of protocol data units. However, due to the limited frame size of 8 byte in classical CAN, SecOC is of limited use on classical CAN bus [146]. As a result of this, on the not yet adopted authentication of ECUs, it is possible to launch impersonation attacks, where one ECU – which is already controlled by an attacker – sends messages that utilize an ID of another ECU. Most current research approaches on the detection of such ECU impersonating attacks use physical fingerprinting by voltage or timing analysis with specific hardware [147] [148].

With respect to category 3 (packet insertions), where malicious packages are inserted outside their usual frequency [149] implemented a method using *One-Class Support Vector Machines*

(OCSVM) and [150] proposed a lightweight intrusion detection system based on small time differences, while [151] utilize LSTM.

With respect to category 4 (sequence context anomalies), in case the attacker acts within the given specifications and does not attract attention by obvious frequency manipulation, the state-of-art methods comprise OCSVM [152], neural networks [153], hidden Markov models [154] [145], process mining [155] or time series analysis [156], Hamming distance between payloads of two consecutive messages with same CAN ID [157], transition matrix [158] for valid ID sequences, and characteristic functions [159] focused on the validity of payload values and changes in said values.

Comparisons of different Machine Learning (ML) algorithms for intrusion detection within CAN data are given in [160] and [161]. LSTM, Gated Recurrent Units (GRU) and Markov models are used in [160], while OCSVM, SVM, sequential neural networks and LSTM are used in [161]. Surveys on intrusion detection systems for in-vehicle networks are provided in [162] and [163].

Because the content structure (see Figure 15) and semantics of the CAN payload is usually kept secret, most methods mentioned above view the payload as unstructured sequence of 8 bytes [156] or 64 bits.

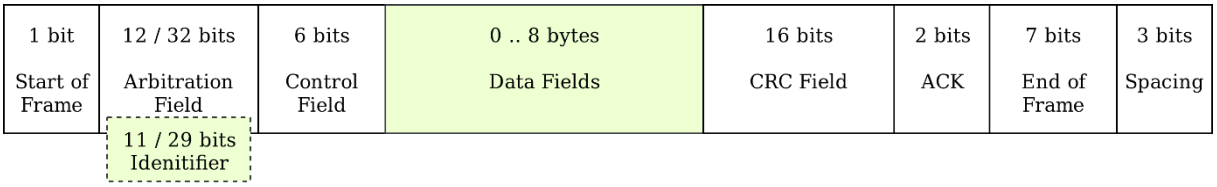

Figure 15 Structure of the CAN message

6.1.2. Proposed Approach/Technology

We now describe the parts of the E-CORRIDOR analytics architecture (see Figure 16) to be used for the intrusion detection in vehicular networks. The proposed components comprise the following functionalities:

- Classification at edge (in-vehicle)
- Anomaly reporting (in-vehicle)
- Classification at back-end (E-CORRIDOR platform)
- Anomaly reporting (edge to back-end)
- Model creation (learning in back-end)
- Model deployment (transfer model back-end to edge)
- Continuous model improvement

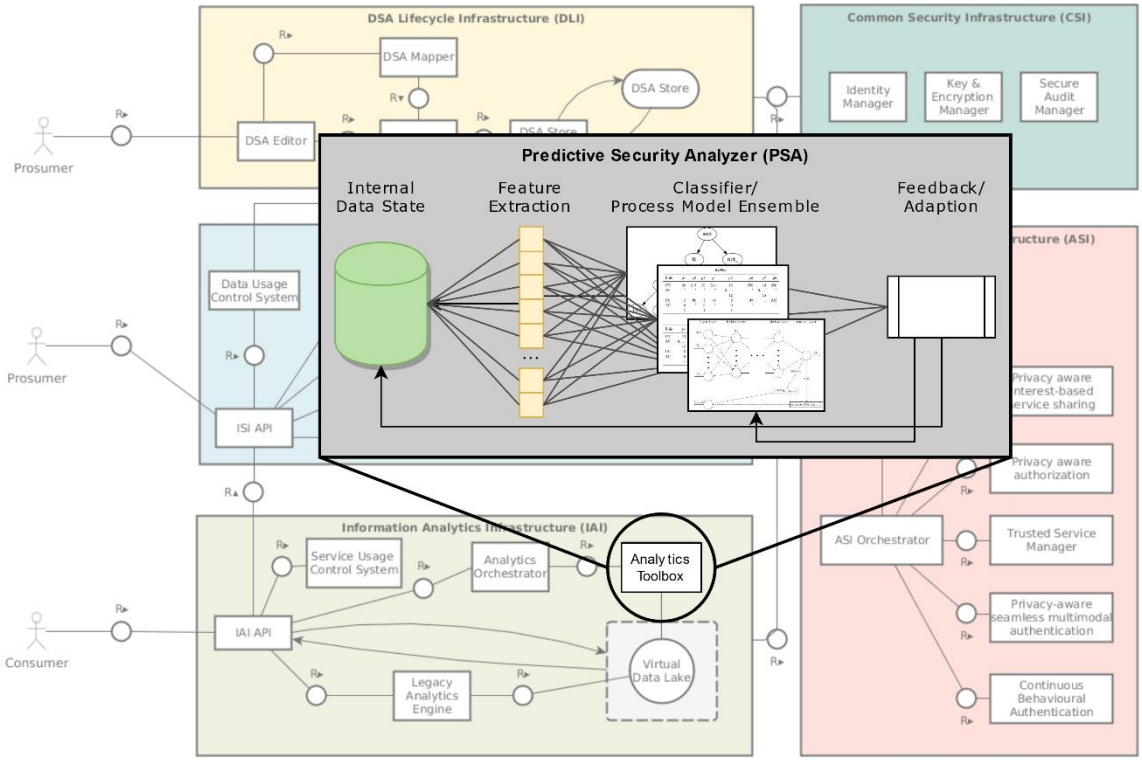

Figure 16 The Automotive IDS component in the analytics toolbox of the IAI

The basis of the automotive IDS available in the analytics toolbox will be constituted by a novel model-based approach for predictive security analysis at runtime. It is based on the Predictive Security Analyzer (PSA) developed in the FP7 project MASSIF. The PSA observes the operation of a managed system by analyzing its events. For the application in the E-CORRIDOR project, several enhancements and adaptations will be implemented, in particular by taking into account the needs of the security services in multimodal transport.

For example, the accuracy of the payload structure model in CAN bus data heavily influences the accuracy of anomaly detection models. Therefore we evaluate this influence with respect to the E-CORRIDOR sensor value structure on the results of different intrusion detection methods. We analyze if an improved alignment is helpful to detect anomalies introduced by complex, hidden intrusions.

In order to cover conceptually different modeling and reasoning techniques, we adapted an artificial neural network approach as well as a characteristic functions based intrusion detection approach to utilize such message streams on the CAN bus. For this we developed a set of test vectors based on log files of a vehicle enriched by different intrusions based on real-life scenarios. We have injected simulations of intrusions which mask certain sensor values within the respective messages. The effectiveness of the developed methods has been demonstrated in various experiments [164].

6.1.3. Data Format Requirement

To reliably evaluate different automotive intrusion detection mechanisms, we utilize a multitude of public CAN bus datasets, such as [165] and [166], as well as our own dataset [159]. In order to successfully test a dataset, several requirements must be met. For one, we require information on which ECU within the vehicle is responsible for sending the data on the bus,

this is usually communicated through its ID and identification of the respective bus. Additionally, the exact timing of the messages must have been recorded along with the complete payload of the message, as well as the respective length of the payload. To improve classification, we also utilize additional information of the vehicles' bus system, for example, the exact payload structure or list of valid IDs on the bus. All public datasets used for evaluation and testing of the automotive intrusion detection approaches contain previously recorded or introduced intrusion messages for several different attack scenarios. To prevent misclassifications or other inaccuracies, all attacks and their respective introduced messages must be tagged in some way. Either complete metadata on the intrusion must be provided, where the exact beginning and ending of the intrusion as well as the structure of the introduced messages is described or the individual messages must be labelled, to mark intrusion messages individually. There is no requirement on the actual structure of the dataset other than those previously mentioned. To evaluate our approaches, we can extract the required information from any kind of log format provided, if the information is present. In Appendix A.2.1 Examples of CAN bus Datasets we show excerpts of different public CAN bus data sets and how the required information is logged there.

After the classification of messages and the potential detection of an anomaly within the vehicles data, the results need to be distributed to different components of the E-CORRIDOR system for visualization or further processing. For this we have decided to use the Structured Threat Information eXpression (STIX) (cf. https://oasis-open.github.io/cti-documentation/) as the intrusion reporting format within the E-CORRIDOR platform. In Appendix A.2.2 Examples of Alert Indicators in STIX Data Format we show by examples how alert information can be represented in STIX format.

6.1.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-CANIDS-01	MUST	Emulate a working ECU inside the in-vehicle network to generate, collect, share and analyze CAN bus data for S2C-UC-06, ISAC-UC-02, ISAC-UC-07.	• E-CORRIDOR-Tst-ISAC-01 • E-CORRIDOR Ope-05
E-CORRIDOR-IAI-CANIDS-02	SHOULD	Support device that is compatible with OBD (or CAN BUS) for monitoring and sending GPS and driving behavior data.	• E-CORRIDOR-Tst-S2C-02
E-CORRIDOR-IAI-CANIDS-03	SHOULD	Support Pilot ISAC Test Bed & Production Requirements	• E-CORRIDOR-Tst-ISAC-01, • E-CORRIDOR-Tst-ISAC-02, • E-CORRIDOR-Tst-ISAC-03, • E-CORRIDOR-Tst-ISAC-04
E-CORRIDOR-IAI-CANIDS-04	SHOULD	Support an intrusion protection system able to authenticate the ECU in an intra-vehicle network when it aims at sending cross partition CAN frame	• E-CORRIDOR-Tst-Int-ISAC-02
E-CORRIDOR-IAI-CANIDS-05	MUST	CAN IDS must work at the edge	• E-CORRIDOR-Ope-02 edge
E-CORRIDOR-IAI-CANIDS-06	COULD	CAN IDS could support deployment in cloud and edge or collaboratively in the cloud	• E-CORRIDOR-Ope-01 (both) • E-CORRIDOR-Ope_03 (collaboratively in the cloud)
E-CORRIDOR-IAI-CANIDS-07	SHOULD	CAN IDS should support intrusion detection reporting E-CORRIDOR cloud by means required by respective use cases.	• E-CORRIDOR-DA-06 • E-CORRIDOR-DS-19 (push)

6.1.5. Application to Pilots

Pilot	S2C and ISAC pilots
Reference to Use cases or User stories	• S2C-US-06 Cybersecurity notifications: communicate about threats • ISAC-US-06 Automotive cyber-threat information analysis • ISAC-US-07 sharing automotive cyber-threat information analysis
Brief description of the Use cases or User stories	The above use cases refer to the automotive cyber-thread information analysis through intrusion detection mechanisms, as well as the sharing of automotive cyber-threat analysis results with different components of the E-CORRIDOR platform and thereby enabling the system to notify and communicate about threats.
Match of the proposed approach/technology with the USs/UCs	The proposed technology identifies anomalous behavior possibly related to attacks on the vehicle and reports them to the E-CORRIDOR platform.

6.1.6. Potential Synergies

Synergies with other components - Work package and Task	• T7.1 Secure routing for driver identification – Driver DNA • T7.5 EARNEST CAN-IPS
Title/brief description of the task	Data Analytics for Driver and Passenger Identification in Task 7.1 analyze sensor data to perform driver and passenger identification. Sensor data are collected from cars such as CAN bus messages. Machine learning algorithms that create models for driver and passenger used for identification could possibly use similar algorithms as CAN IDS. EARNEST is an Intrusion Protection System (IPS) to prevent unauthorized ECUs to send possible malicious CAN frames on the bus.
Description of the potential synergy with risks and opportunities	CAN IDS could support in particular the driver identification in Task 7.1 because in-vehicle data on CAN bus could be exploited for behavioral and driving style analysis. CAN IDS can be used to trigger EARNEST CAN IPS developed in task 7.5 by specific alerts. Thus, CAN IDS analytics can support and extend the protection of malicious activities within a vehicle.
Dependencies on other components	none

6.2. Fully Homomorphic Encryption-based intrusion detection [E-CORRIDOR-IAI-FHEIDS]

An Intrusion Detection System (IDS) is a software application that scans a network or a system and raises alerts in presence of anomalous, harmful or policy breaching activities. Any malicious or violation event is reported either to an administrator or collected centrally using a Security Information and Event Management (SIEM) system. A SIEM system integrates outputs from multiple sources and uses alarm filtering techniques to differentiate malicious activities from false (positive) alarms. We propose a Network IPS monitoring the connection with blacklisted IP addresses. Notably, all the stored IP addresses are in homomorphic encryption format in order to respect security and privacy constraints.

6.2.1. State of the Art

Please refer to Section 4.1.1

6.2.2. Proposed Approach/Technology

In the E-CORRIDOR project and due to Pilots requirements, we propose FHE analytics services processing on two data types:

- IPv4 addresses;
- ASCII strings for checking spam with specific patterns

We will implement different FHE analytics services which will allow the following operations:

- Testing if two encrypted IPv4 addresses are equal or not;
- Testing if an encrypted IPv4 belongs to a list of encrypted IPv4 addresses (e.g. to check if the IP is in a list of malicious IPs);
- Testing if an encrypted pattern appears in encrypted text;
- Testing if the *maxlen* first letters of an encrypted string belongs to a list of encrypted strings, where strings can have variable length and where *maxlen* is a parameter of integer type (e.g., to check if a hostname is in a list malicious hostname, if a username is in a list of sensitive accounts).

The analytics operations will be performed with CEA's FHE technologies. FHE computation services on (homomorphic) ciphertexts, including the FHE analytics services proposed above, can all be decomposed on the elementary operations of homomorphic additions and homomorphic multiplications over the input bits of data.

Effectively, Homomorphic Computation enables an untrusted server to evaluate arithmetic circuits on ciphertexts without being able to decrypt inputs and outputs. In concrete terms, it is used to evaluate polynomials over encrypted bits.

As a first illustrative example, we can consider a polynomial P(X,Y)=X+Y over integers and two integers 3 and 5. On clear data, the evaluation of the polynomial returns 8. In homomorphic cryptography, we manipulate encrypted data. By denoting with HE a homomorphic encryption function, in this case, we have two encryptions HE(3) and HE(5) of the integers. The result of the evaluation of P over these ciphertexts is HE(8). That is an encryption of the expected result.

These polynomials can be multivariate and described with Boolean circuits. These circuits can be described by using two Booleans gates: AND gates (binary multiplication) and XOR (binary addition). On such a representation, there are two important parameters that have to be

minimized: (i) number of AND gates and (ii) circuit multiplicative depth i.e., the maximal number of AND gates between an input and an output of the circuit.

As a second more complex illustrative example, we consider the polynomial Q(X,Y)=X^2*Y^2. We can evaluate this polynomial in different ways depending on operation order. A circuit permits to indicate the computation order. Let us take a first circuit representing how Q(X,Y) is computed:

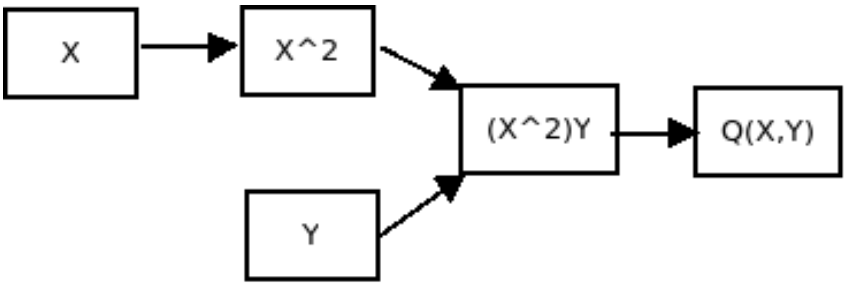

Figure 17: A Boolean circuit – un-optimized

The left boxes represent the circuit inputs. The right box is the circuit output. With this representation, the multiplicative depth is the maximal number of arrows between an input and an output of the circuit. Here, it is 3. We can do better (that is minimising the multiplicative depth) by changing the order of computations:

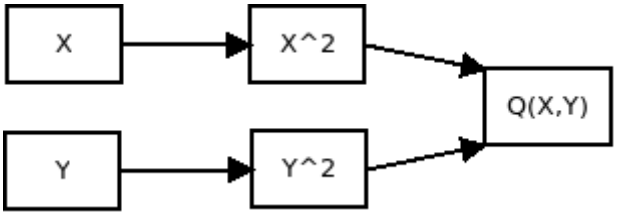

Figure 18: A Boolean circuit – optimised

In this manner, the multiplicative depth is minimized to 2. This permits to decrease time and memory requirements.

Ideally, the multiplicative depth should be less than 20. Time and memory needs mainly depend on security level and multiplicative depth. There is no standardised homomorphic cryptosystem. Earlier, most of the homomorphic cryptosystem proposed only one homomorphic operation (addition or multiplication for instance) and thus less applications were proposed. Note that a homomorphic scheme is probabilistic, so an attacker which has only access to ciphertexts could decrypt it if they have secret key.

The "pattern searching" algorithm of our IDS receives in input a list of blacklisted IP addresses stored in a database in encrypted format. An IP address under analysis, represented as a vector of its encrypted octets, is subtracted to the addresses stored in the database. If the encrypted result is a zero vector then it means that the given address matches the one in the database. In practice, pattern searching uses several techniques to reduce the size of the output cypher-texts and improve performance. Such an optimization will constitute the effort for the next months of the project.

6.2.3. Data Format Requirement

CEA's FHE technology tool needs constant data length as input, which requires a precomputation for certain data types. Indeed, IPv4 addresses satisfy this requirement because they are represented by 4 bytes, whereas strings do not and therefore their length is a parameter to consider.

Strings representations depend on:

- Character encoding;
- String size.

In our implementation, character encoding is (extended) ASCII, where a character is represented with one byte. To address the non-constant string size issue, a solution is to encode the string with zero padding and truncation. The encoded data is then stored as a fixed number of bytes (this number is an additional parameter to consider).

Under those assumptions, the data representation becomes:

- Each IPv4 is stored as 4 bytes;
- Each ASCII character is stored as 1 byte;
- Each ASCII string is stored as X bytes, where X is a parameter;

We choose IPv4 rather than IPv6 addresses and ASCII text format rather than UTF-8 encoding, because they both differ in data representation size (they both use more bytes): in homomorphic cryptography, this parameter can have a significant impact on time and memory requirements.

To sum up, the parameters in our FHE analytics are:

- The input data (IPv4s, String text format);
- Encoded data size (a constant integer for IPv4 addresses, an integer parameter for strings);
- Number of data (the list size is a parameter).

For example, if we would like to analyse the IP addresses in the Common Event Format (CEF) log file and check whether they are blacklisted, then the input of CEF log file would contain:

```
2018-11-27 15:59:19.000 47.000 TCP 146.48.36.2:22  -> 116.31.116.6:21115      22   3973     1
2018-11-27 16:00:03.000 0.000  TCP 146.48.36.2:43497 -> 185.156.177.129:59275  0   0     1
2018-11-27 16:00:12.000 0.000  TCP 146.48.36.2:22 -> 117.33.114.6:45336         1  74     1
2018-11-27 16:00:12.000 0.000  TCP 146.48.36.2:22 ->  119.37.236.1:44929      1   74     1
```

After pre-processing for filtering only the destination IP address, the input file read by the FHE IDS components would be:

```
1; 116.31.116.6
2; 185.156.177.129
3; 117.33.114.6
4; 119.37.236.1
```

6.2.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-FHEIDS-01	MUST	Emulate a working ECU inside the in-vehicle network to generate, collect, share and analyze CTI, connection logs data for S2C-UC-06, ISAC-UC-02, ISAC-UC-07.	• E-CORRIDOR-Tst-ISAC-01 • E-CORRIDOR Ope-05
E-CORRIDOR-IAI-FHEIDS-02	SHOULD	Support device that is compatible with OBD (or CAN BUS) for monitoring and sending GPS and connection behavior data.	• E-CORRIDOR-Tst-S2C-02
E-CORRIDOR-IAI-FHEIDS-03	SHOULD	Support Pilot ISAC Test Bed & Production Requirements	• E-CORRIDOR-Tst-ISAC-01, • E-CORRIDOR-Tst-ISAC-02, • E-CORRIDOR-Tst-ISAC-03, • E-CORRIDOR-Tst-ISAC-04
E-CORRIDOR-IAI- FHEIDS -04	SHOULD	Support an intrusion protection system able to identify IP in black list or spam content	• E-CORRIDOR-Tst-Int-ISAC-02
E-CORRIDOR-IAI- FHEIDS -05	MUST	FHE IPS must work at the edge	• E-CORRIDOR-Ope-02 edge
E-CORRIDOR-IAI- FHEIDS -06	COULD	FHE IPS analysis could support deployment in cloud and edge or collaboratively in the cloud	• E-CORRIDOR-Ope-01 (both) • E-CORRIDOR-Ope_03 (collaboratively in the cloud)
E-CORRIDOR-IAI- FHEIDS -07	SHOULD	FHE IPs should support intrusion detection reporting E-CORRIDOR cloud by means required by respective use cases.	• E-CORRIDOR-DA-06 • E-CORRIDOR-DS-19 (push)

6.2.5. Application to Pilots

Pilot	AT, S2C, ISAC
Reference to Use cases or User stories	• AT-US-04: Advanced Security Analytics Services • ISAC-US-03: ISAC-MMT cyber-threat information analysis • ISAC-US-04: ISAC-MMT cyber-threat notification • ISAC-US-07: sharing automotive cyber-threat information analysis • ISAC-US-06: Automotive cyber-threat information analysis • S2C-US-06 Cybersecurity notifications: communicate about threats • S2C-US-07 Secure sensitive data that would be shared from end to end
Brief description of the Use cases or User stories	The use cases involves collection and processing of cyber-security logs.
Match of the proposed approach/technology with the USs/UCs	The FHE-based IDS will process those logs to identify any blacklisted address attempting to connect to the internal network. Similarly, the component will be effective in identifying an infected machine unconsciously trying to connect to remote servers (e.g., a command and control server) if infected by a malware.

6.2.6. Potential Synergies

Synergies with other components - Work package and Task	• T7.2
Title/brief description of the task	OpenAPI for FHE operations
Description of the potential synergy with risks and opportunities	Both the components are based on the same FHE engine but the different applications require special customizations and optimizations.
Dependencies on other components	None

6.3. Intrusion Protection System – Earnest [E-CORRIDOR-IAI-CANIPS]

Modern vehicles are composed by many micro-controllers called Electronic Control Units (ECUs) that are controlled by software. ECUs regulate all the functionalities of a vehicle, including many safety-critical functions (the steering and breaking) and other possible untrusted functionalities (the infotainment system).

ECUs are physically connected to the vehicle network, and communicate one other through specialized protocols, such as the Controller Area Network (CAN) bus protocol. However, the

CAN protocol is not secure by design. The lack of mechanisms to guarantee security of on-board communication in conjunction with network design choice may be cause of cyber-attacks. One of the most famous example is the one of Miller and Valasek to a Jeep Cherokee in 2015 [167] in which the two researchers remotely drove the vehicle. In 2018, the Keen Security Lab presented a set of vulnerabilities of BMW cars that make them prone to remote access [168]. In particular, the researchers injected Unified Diagnostic Services (UDS) frames into the CAN network bypassing the central gateway. Another example is a recent attack on a Toyota Lexus, introduced in March 2020 by the Keen Security Lab [169].

In this interesting panorama, we propose EARNEST as an Intrusion Protection System (IPS) to prevent unauthorized ECU to send possible malicious CAN frames on the bus. The application scenario we consider consists of an active attacker who gets access to the CAN bus network, either exploiting a local or remote connection. The intra-vehicle network is designed in such a way that two or more partitions of the network are put in place to isolate at least the safety-critical functionalities to the untrusted ones. Such an attacker may alter the behavior of a vehicle by obtaining access to the vehicle network, for instance, by compromising an ECU. Gaining the control of the ECU, the attacker could be able to inject customized frames (Fuzzing attack), or to perform Replay attacks. EARNEST prevents both of them. Whenever an ECU sends a frame from a partition to another (cross- partition frame), EARNEST halts the frame and challenges the ECU. The challenge consists in performing a simple operation on agreed dynamically generated frame. If the challenge successfully ends the frame is forwarded, otherwise, discarded.

6.3.1. State of the Art

In academic literature, several solutions were proposed to cope with security issues on the CAN bus. Such solutions can be classified in two main categories: add security to CAN bus, e.g., [170] or Intrusion Detection/Protection Systems e.g., [171]. Here, we discuss EARNEST with existing results about intrusion detection and protection mechanisms based on anomaly detection on CAN communication.

An intrusion detection and prevention mechanism was proposed by Miller and Valasek in [167]. It consists in an anomaly detector device that can be plugged into the vehicles OBDII port. The systems acts by checking anomaly on the communication bus traffic pattern. Once anomalies were identified, they disabled the CAN.

Also the IDS presented in [172] is based on anomaly detection. In particular, it is based on detecting an anomaly in the frequency of the messages that are sent. Once the anomaly is detected, an alert is sent to the infotainment system to inform the driver that a possible intrusion occurred.

In [157], the author proposes an anomaly detection mechanism based on the Hamming Distance [173] between consecutive payloads of CAN frames with the same ID. They do not consider the whole payload but only the observed distance. They provide also performances related to the efficiency of the approach in the detection of Replay and Fuzzing attack. In [158], the same authors present an Intrusion Detection mechanism based on machine learning. The proposed algorithm is able to build a model of the normal behavior of a CAN network based on the recurring patterns within the sequence of message IDs observed in the CAN Bus. In this way the proposed IDS (Intrusion Detection System) is able to address Replay attacks, and the bad or mixed injection of messages on the bus.

For the best of our knowledge, in literature Intrusion Protection Systems on CAN bus are not DBC-challenge based. EARNEST provides a strategy to protect the CAN bus network as soon as a possible attack is in action. This is the main novelty of EARNEST: establishing if the frame content are generated by an attacker using a challenge that allows EARNEST to prove the licit ECU behavior. In fact, by using the challenge mechanism, in which we assume that the DBC is employed as the secret shared among ECU and EARNEST, our approach is able to cope with Replay and Fuzzing attacks.

6.3.2. Proposed Approach/Technology

Seen the continuous evolution that the automotive field is facing, without considering a specific vehicle network, we focuses on the last version of in-vehicle network where partitions are present. In this scenario, the diagnostic socket, such as the OBDII [174] port, is also considered as separate partition connected to the others through a central gateway (CGW) where EARNEST is installed. The goal of EARNEST is to monitor those CAN frames, named cross-partition frames that are generated from a partition and addressed to other ones. We take also into account attacks that may come from the diagnostic bus. A common replay attack that can be executed here is represented by the sniffing of diagnostic frames that were created for instance by a diagnostic tool. Those frames can be easily replicated by any device capable to inject CAN frames into the bus. Due to the lack of any security mechanisms, e.g., packet authentication or frame freshness, the addressed ECUs will receive the frames and execute the corresponding actions.

Considering the above cases, EARNEST aims at blocking attacks by verifying the sender authenticity by means of a challenge method. In particular, EARNEST's goal is to stop any kind of replay and fuzzing attack that may be generated by ECUs when the frames are cross-partition ones. The security of EARNEST is based on the following points:

- DBC enables ECUs of a specific vehicle to correctly generate and interpret messages' payload and translate them into signals that carry out the expected functionality. DBC is proprietary of OEM and it can be considered as a long-term secret only known by EARNEST and legitimate ECUs.
- The challenge set represents the type of challenges that EARNEST will ask the ECU that is trying to send a cross-partition frame.
- The encoding generation method defines how frames, needed for the challenge, must be generated by the ECU that sent a cross-partition frame.

EARNEST adopts a challenge-based approach that starts once an ECU sends a cross-partition frame. EARNEST challenges the ECU asking to resolve a challenge as part of the handshake protocol. The ECU has to solve the challenge and sends back the correct answer to EARNEST. The goal is twofold: i) it allows the authentication of the sender ECU, since only legitimate ECUs knows the DBC, the set of challenges, the encoding method, and ii) the broadcast of the original frame to a cross-partition that may come from untrusted zone. In case of incorrect answer, the frame is discarded.

6.3.3. Data Format Requirement

EARNEST works using the CAN bus protocols. ECUs that generate cross-partition frames are challenged by EARNEST and failed challenges are reported in STIX format. Examples of STIX

objects generated by EARNEST are reported, for the interested reader, in Appendix A.2.3 Examples of EARNEST Reports in STIX Data Format.

6.3.4. Platform Requirements

ID	Priority	Requirement	In order to fulfil Platform Requirement(s)
E-CORRIDOR-IAI-CANIPS-01	MUST	Emulate a working ECU inside the in-vehicle network to generate, collect, share and analyze CAN bus data for S2C-UC-06, ISAC-UC-02, ISAC-UC-07.	• E-CORRIDOR-Tst-ISAC-01 • E-CORRIDOR Ope-05
E-CORRIDOR-IAI-CANIPS-02	SHOULD	Support device that is compatible with OBD (or CAN BUS) for monitoring and sending GPS and driving behavior data.	• E-CORRIDOR-Tst-S2C-02
E-CORRIDOR-IAI-CANIPS-03	SHOULD	Support Pilot ISAC Test Bed & Production Requirements	• E-CORRIDOR-Tst-ISAC-01, • E-CORRIDOR-Tst-ISAC-02, • E-CORRIDOR-Tst-ISAC-03, • E-CORRIDOR-Tst-ISAC-04
E-CORRIDOR-IAI-CANIPS-04	SHOULD	Support an intrusion protection system able to authenticate the ECU in an intra-vehicle network when it aims at sending cross partition CAN frame	• E-CORRIDOR-Tst-Int-ISAC-02
E-CORRIDOR-IAI-CANIPS-05	MUST	CAN IPS must work at the edge	• E-CORRIDOR-Ope-02 edge
E-CORRIDOR-IAI-CANIPS-06	COULD	CAN IPS could support deployment in cloud and edge or collaboratively in the cloud	• E-CORRIDOR-Ope-01 (both) • E-CORRIDOR-Ope_03 (collaboratively in the cloud)
E-CORRIDOR-IAI-CANIPS-07	SHOULD	CAN IPS should support intrusion detection reporting E-CORRIDOR cloud by means required by respective use cases.	• E-CORRIDOR-DA-06 • E-CORRIDOR-DS-19 (push)

6.3.5. Application to Pilots

Pilot	ISAC and S2C
Reference to Use cases or User stories	• S2C-UC-06: Security analytics: Notifications and threat/attack management • ISAC-US-01: Public cyber-threat information collection • ISAC-US-02: Private transportation sector data collection
Brief description of the Use cases or User stories	The above use cases refer to the collection of data and the elaboration of such data in a privacy preserving way
Match of the proposed approach/technology with the USs/UCs	This analytics will help the stakeholders to discover and report malicious activities related to in-vehicle protocol network such as the CAN bus protocol.

6.3.6. Potential Synergies

Synergies with other components - Work package and Task	• T7.5 Automotive Intrusion Detection (CAN bus IDS)
Title/brief description of the task	The above tasks refer to automotive intrusion detection for
Description of the potential synergy with risks and opportunities	This analytics can support and extend the protection of malicious activities within a vehicle. This will be done in collaboration with the analytics introduced in Section 6.1.
Dependencies on other components	Automotive Intrusion Detection

7. Pilot specific analytics

The ISAC offers pilot-specific analytics based on the data gathered from public sources (security databases, online information) and the results produced by the security analytics provided by the other pilots. The analytics toolbox contained in the IAI component instantiated in the ISAC offers two types of analytics. The cyber data label assignment and the cyber data visualization tool. Both of them are integrated into the ISAC in order to extract high-level correlation between data collected and make it possible an intuitive visualization of the aggregation. Figure 19 shows the list of analytics contained in the IAI component of the ISAC.

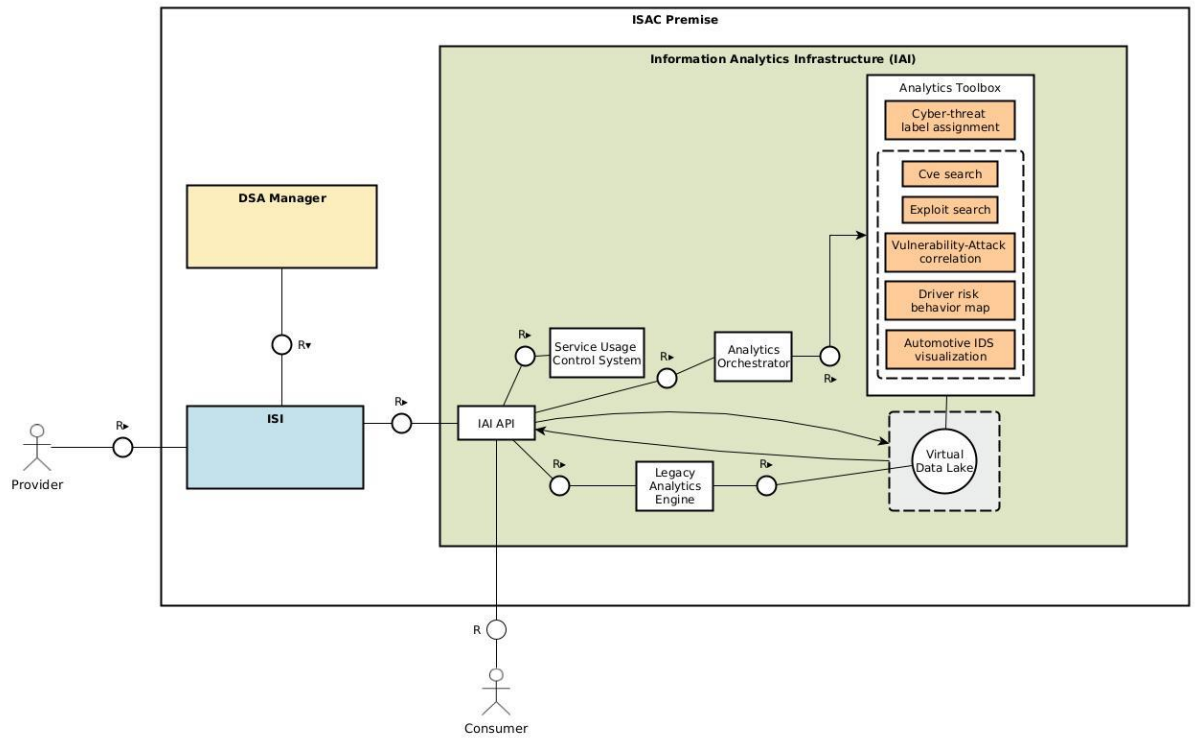

Figure 19 Pilot-specific analytics of the ISAC pilot

7.1. Cyber data label assignment [E-CORRIDOR-IAI-CDLA]

This analytics is exploited as the main element for the ISAC cyber-threat notification workflow. The ISAC is bound to collect a huge amount of information from various providers, and at the same time, it must redistribute the collected data timely to interested stakeholders. Selecting the right information to be sent to the right partner is of high importance. In fact, failing to provide information on a relevant vulnerability might increase the exposure time of the stakeholder to cyber-attacks. On the other hand, flooding the stakeholder with non-pertinent information increases the processing times and might result in a lowered attention to actual security threats. This analytics exploits text analysis using Natural Language Processing techniques and clustering to separate the received and computed information into separate topics to which the stakeholders can subscribe. In such a way, they will be automatically notified of the information that is relevant to them.

7.2. Cyber data visualization [E-CORRIDOR-IAI-CDV]

The cyber data visualization analytics is a macro analytics used to present and speed up the information retrieval process. It is composed of a set of visualization analytics described in the following.

- *Security report*: The security report is a set of statistical analyses extracted from the entire data collection available in the ISAC. It is regularly updated every time new cyber data is collected, and it is offered to the consumer to increase awareness in the cyber-threat field. The list of the reporting data is described in Table 3.

Table 3 Data in the security report of the ISAC

Data	Description
Vulnerabilities	Distribution of the number of CVE in the last 10 years.
	Distribution of the degree of danger (CVSS) of all CVEs discovered.
	Distribution of the number of CVEs concerning the most used transportation software/hardware.
	Distribution of the degree of hazard of the CVE discovered in the current year.
	Latest CVE discovered
Exploit	Distribution of the number of exploit types in the last 10 years.
	Distribution of the number of exploits discovered in the last 10 years.
	Distribution of the number of exploits related to the most common transportation software.
	Latest exploits discovered

- *CVE search*: This service offers the possibility of searching public domain information related to known security hardware and software vulnerabilities. This service provides a general description of the vulnerabilities reporting the publishing date, a short description, the CVE score, and the impact on integrity, confidentiality, and availability properties. The consumer can research the vulnerabilities specifying an interval time or a specific keyword. Data exploited: Public vulnerability database collected from NVD (*https://nvd.nist.gov*).
- *Exploit search*: This analytics offers the possibility of searching information about the exploits performing research by date or keyword. This analytics shows the date, description, and the specific platform on which the exploit is effective. Data exploited: Public exploits information collected from exploit-db (https://www.exploit-db.com/).

- *Vulnerability attack correlation*: This analysis allows the consumer to explore the interconnection between the vulnerabilities and attack patterns of specific known software or hardware. Additionally, the analytics provides recommendations for attacks and vulnerability mitigation.
- *Driver risk behavior map*: The analytics exploits the data provided by the results of the Driver DNA analytics (E-CORRIDOR-IAI-SR). As explained in Section 2.1, the analytics provides the driver risk profile, e.g., more aggressive, speeding more frequently. Correlating these results with the geolocalization of the vehicles, the ISAC analytics can produce a map of the city reporting the average driver behavior in each road section. In such a way, it is possible to highlight the most dangerous section of the city and increase the drivers' awareness.
- *Automotive intrusion detection visualization*: This analytics is based on the automotive intrusion detection system analysis results explained in Section 6.1 (E-CORRIDOR-IAI-CANIDS). The produced result is the classification of the CAN bus messages and the potential detection of an anomaly within the vehicle's data. Such information shared by multiple vehicles with the ISAC is exploited to create a graphic visualization of the intrusion information, incident, related types, and incident classes of a single vehicle or create a visualization of the correlation of the incident detected in different vehicles.

8. Map of Data Analytics Techniques to Pilot Requirements

All the analytics discussed in this document have been designed by taking into account requirements. This is remarked in Table 4 where, for the use cases of each of the three project pilots (as defined in D2.1, D3.1 and D4.1), the analytics contributing to their realization are specified. It is worth remarking that the analytics generally need the joint work with the other subsystems of the E-CORRIDOR framework (e.g., DLI, ISI or ASI). Where no analytics is involved in a use case the corresponding note is reported in the table. Pilots' requirements are expressed along with their priority by following the MoSCoW (Must have, Should have, Could have, and Won't have but would like) technique [175]. In case the "application to pilots" tables of the analytics components have been described with respect to user stories, the related use cases are reported here for the sake of homogeneity and conciseness.

Table 4 Match of pilots' use cases with the analytics in the toolbox of the E-CORRIDOR framework

Pilot	Use Case ID	Use Case Name	Priority	Analytics Identifier	Name of the analytics
Airport- train (AT)	AT-UC-01	PRM Passenger Assistance and Authorization	Must	E-CORRIDOR-IAI-PL E-CORRIDOR-IAI-PBI E-CORRIDOR-IAI-FR E-CORRIDOR-IAI-AR	Passenger location and flow optimization Passenger: Identification, Behavior, Context Face recognition – passenger authentication Activity recognition – passenger authentication
	AT-UC-02	Passenger and Baggage Contextual Identification	Must	E-CORRIDOR-IAI-PBI E-CORRIDOR-IAI-AR	Passenger: Identification, Behavior, Context Activity recognition – passenger authentication
	AT-UC-03	Contactless Passenger Authentication and Authorization	Must	E-CORRIDOR-IAI-PBI E-CORRIDOR-IAI-GA E-CORRIDOR-IAI-FR E-CORRIDOR-IAI-AR	Passenger: Identification, Behavior, Context Gait analysis – passenger authentication Activity recognition – passenger authentication
	AT-UC-04	Privacy-preserving Passenger Monitoring	Should	E-CORRIDOR-IAI-PL	Passenger location and flow optimization

				E-CORRIDOR-IAI-PBI E-CORRIDOR-IAI-AR	Passenger: Identification, Behavior, Context Activity recognition – passenger authentication
	AT-UC-05	Passenger Analysis Opt-In Opt-Out	Must	* the DSA available in the DLI are expressed in a human intelligible format so that the passenger can accept the way data are shared in the ISI and used by other prosumers	
	AT-UC-06	Single Sign-On Authentication	Must	E-CORRIDOR-IAI-FR E-CORRIDOR-IAI-AR	Face recognition – passenger authentication Activity recognition – passenger authentication
	AT-UC-07	Multi-Modal Ticketing	Could	* the passenger's information are stored in the eWallet located in the ISI and the federated authentication component of the ASI provides the required operations	
	AT-UC-08	Service Access Through Bring Your Own Device	Could	E-CORRIODR-IAI-GA	Gait analysis – passenger authentication
	AT-UC-09	Sharing of Service Access Data	Must	E-CORRIDOR-IAI-PL	Passenger location and flow optimization
	AT-UC-10	Run Collective Security Analytics	Could	E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-FHEIDS	OpenAPI for Fully Homomoprhic Encryption Fully Homomorphic Encryption-based intrusion detection
	AT-UC-11	Notification of Service Disruption	Could	E-CORRIDOR-IAI-PL	Passenger location and flow optimization
	AT-UC-12	Passenger Flow Overview and Prediction	Should	E-CORRIDOR-IAI-PL E-CORRIDOR-IAI-PBI E-CORRIDOR-IAI-AR	Passenger location and flow optimization Passenger: Identification, Behavior, Context Activity recognition – passenger authentication
	AT-UC-13	Privacy-aware Behavioral Identification	Should	E-CORRIDOR-IAI-PBI	Passenger: Identification, Behavior, Context

				E-CORRIDOR-IAI-GA E-CORRIDOR-IAI-AR	Gait analysis – passenger authentication Activity recognition – passenger authentication
	AT-UC-14	Notification on PRM Passengers Location	Could	E-CORRIDOR-IAI-PL	Passenger location and flow optimization
Smart cities and car sharing (S2C)	S2C-UC-01	eWallet: Sign in/Log in	Must	E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-FHEIDS	OpenAPI for Fully Homomoprhic Encryption Fully Homomorphic Encryption-based intrusion detection
	S2C-UC-02	Socio-geographic dependent micro-subsidies	Must	* this use case requires the presence of the passenger's eWallet stored in ISI and the trusted service manager component of the ASI	
	S2C-UC-03	Trip planning and carbon footprint analysis	Could	E-CORRIDOR-IAI-MMIP E-CORRIDOR-IAI-CFA	CO2-aware Trip Planning Carbon footprint analysis
	S2C-UC-04	Sharing service data with Transport authority	Could	E-CORRIDOR-IAI-FHEC	OpenAPI for Fully Homomoprhic Encryption
	S2C-UC-05	Informing travelers about data usage and privacy	Could	E-CORRIDOR-IAI-FHEC	OpenAPI for Fully Homomoprhic Encryption
	S2C-UC-06	Security analytics: Notifications and threat/attack management	Should	E-CORRIDOR-IAI-CANIDS E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-FHEIDS	Automotive Intrusion Detection OpenAPI for Fully Homomoprhic Encryption Fully Homomorphic Encryption-based intrusion detection
	S2C-UC-07	Security analytics: Privacy aware interest-based service sharing	Could	E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-FHEIDS	OpenAPI for Fully Homomoprhic Encryption Fully Homomorphic Encryption-based intrusion detection
	S2C-UC-08	Driving behavior recognition	Could	E-CORRIDOR-IAI-SR E-CORRIDOR-IAI-MPCSR	Driver DNA Private Secure Routine

Multi-Modal Transportation Information Sharing and Analysis Center (ISAC)	ISAC-UC-01	Public Cyber-Threat Information (CTI) data collection	Must	E-CORRIDOR-IAI-SR E-CORRIDOR-IAI-MPCSR E-CORRIDOR-IAI-IPS E-CORRIDOR-IAI-CDLA	Driver DNA Private Secure Routine Intrusion Protection System – EARNEST Cyber data label assignment
	ISAC-UC-02	ISAC-MMT sharing data	Must	E-CORRIDOR-IAI-CANIDS E-CORRIDOR-IAI-MMIP E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-CANID	Automotive Intrusion Detection Multi-modal itinerary planning OpenAPI for Fully Homomoprhic Encryption Automotive Intrusion Detection
	ISAC-UC-03	Data sharing agreement	Must	E-CORRIDOR-IAI-MMIP	E-CORRIDOR-IAI-MMIP: Multi-modal itinerary planning
	ISAC-UC-04	Run ISAC-MMT security analysis	Must	E-CORRIDOR-IAI-FHEC E-CORRIDOR-IAI-FHEIDS	OpenAPI for Fully Homomoprhic Encryption Fully Homomorphic Encryption-based intrusion detection
	ISAC-UC-05	Cyber-threat notification	Must	E-CORRIDOR-IAI-CANIDS E-CORRIDOR-IAI-FHEIDS E-CORRIDOR-IAI-CDV	Automotive Intrusion Detection Fully Homomorphic Encryption-based intrusion detection Cyber data visualization
	ISAC-UC-06	Specific transportation sector data collection	Must	E-CORRIDOR-IAI-IPS E-CORRIDOR-IAI-CDV	Intrusion Protection System – EARNEST Cyber data visualization
	ISAC-UC-07	Run local analytics	Must	E-CORRIDOR-IAI-CANIDS	Automotive Intrusion Detection
	ISAC-UC-08	CTI visualization	Must	E-CORRIDOR-IAI-CDV	Cyber data visualization

As can be observed from the above table, the analytics proposed in the toolbox aims at satisfying the requirements expressed by the three project pilots as use cases and thus set the foundations for their further evaluation and maturation in the realistic environments proposed by the pilots, the development of products and their potential adoption.

9. Contributions to the E-CORRIDOR objectives

The E-CORRIDOR framework aims at providing a collaborative, privacy-aware and edge-enabled platform for information sharing, data analysis and security services to the multi-modal transportation domain.

This deliverable is oriented at describing the data analytics and security services offered by the framework according to pilots' requirements and needs specified at M12. In particular, pilots have expressed very heterogeneous requirements with respect to data sources, available sensors, deployment models and system capabilities. On the other hand, it was possible to cluster their needs toward privacy-aware analytics for user identification and authentication, itinerary planning and cyber-security services for transportation entities.

Among the project objectives described in the DoA of the E-CORRIDOR project, the analytics discussed in this deliverable contribute to the following ones (reported here for the sake of completeness):

- Objective 2: E-CORRIDOR will define edge enabled data analytics and prediction services in a collaborative, distributed and confidential way;
- Objective 3: E-CORRIDOR will define a secure and robust platform in holistic manner to keep the communication platform safe from cyber-attacks and ensure service continuity;
- Objective 4: E-CORRIDOR will improve, mature and integrate several existing tools provided by E-CORRIDOR partners and will tailor those to the specific needs of the E-CORRIDOR platform and Pilots;
- Objective 5: E-CORRIDOR will provide mechanisms for seamless access to multimodal transport;
- Objective 6: the framework and the services developed will be used to deliver three pilot products;

Stakeholders in the multi-modal transportation domains have the possibility to collaboratively share data in a privacy preserving manner thanks to the ISI and DLI (Data sharing agreement lifecycle infrastructure) subsystems of the E-CORRIDOR framework. Such data bundles contain highly sensitive (e.g., passport or driving license, biometrics) and confidential information (e.g., pertaining to the functioning of the transportation infrastructure or the cars). These can be analyzed with the privacy-aware analytics provided in the analytics toolbox in a collaborative fashion, thanks to the edge capabilities of the platform. The DSA specified in the DLI are able to control the access to the data and even to the set of analytics deemed trusted by the data producer. These capabilities of the E-CORRIDOR framework are then able to satisfy performance and regulatory constraints requested by the pilots, other than contributing to the *Objective 2* of the E-CORRIDOR project.

In the E-CORRIDOR framework the ASI subsystem is in charge of providing advanced trusted services, authentication and authorization capabilities based also on multi-biometric, multi-factor and behavioral analysis to ensure security for the platform and its users. The analytics presented here other than providing some identity, privacy and security services (see components described in Sections 2, 4 and 6) are, in some cases, also used as building blocks for more complex cyber-security services performed by the ASI. Synergies presented along

each component highlights the latter possibility. Moreover, the intrusion prevention and detection services (see Section 6) directly provide a mean to strengthen the cyber-security and increase the robustness of the pilot infrastructure in line with the goals expressed by the *Objective 3*.

As already remarked, all the analytics in the toolbox have been designed and tailored to the needs of the project pilots, and are actually able to cover a good spectrum of their use cases (see Section 8). The pluggable and privacy-aware capabilities of the toolbox will ensure that further requirements, defined in the next months of the project and even after its completion, can be easily considered and integrated in the framework without any additional effort. Older components not originally compliant with the specifications of the IAI subsystem can be integrated in the legacy analytics engine. All this will ease the integration of existing and new tools in the E-CORRIDOR framework. *Objective 4* is fulfilled by considering that analytics tools are also expected to improve and maturate (up to a TRL 6 or 7) thanks to the closer interaction with experts in the multi-modal transportation domain (represented by the project pilots) and their interest in evaluating the capabilities and features offered by the analytics to satisfy the described use cases.

The goal of a seamless access to the multimodal transportation services expressed in the *Objective 5* of the project is supported in particular by the IAI and ASI subsystems of the E-CORRIDOR framework. In particular the IAI offers a set of passenger and driver identification and authentication capabilities targeting the peculiarity of the project pilots in terms of available sensor data, privacy requirements, and deployment and resource restrictions. The ASI will extend those services with behavioral authentication and authorization, trusted service manager. A frictionless experience will be experience by the user thanks to the exploitation of a multitude of sensors and environmental analysis (please note the different data formats and sources considered by the analytics in the toolbox) able to ensure a seamless and privacy-aware authentication.

The closer collaboration of the technology providers of the project with the pilots and the suitability of the data analytics tools for the use cases of the latter has been expressed, for each analytics component, by the tables summarizing the “application to pilots” (and summarized in Section 8). As a matter of fact, the project pilots are designing themselves the products suitable for their infrastructure (*Objective 6*) through detailed requirements and domain constraints. Given heterogeneity and representativeness of the pilots, security and passenger experience in a pan-European multi-modal transportation can be improved to achieve a really integrated journey when the technology of the E-CORRIDOR framework and its components will further maturate and be ready to be adopted by the pilots and the represented transportation sectors.

Please refer to the corresponding sections of the other project deliverables and in particular to the one in D5.2 for an overview on how the E-CORRIDOR framework contributes to the project goals.

10. Conclusions

This deliverable presented the list of E-CORRIDOR analytics components identified at M12 as relevant to successfully achieve the requirements identified by the project pilots. The latter express a representative cross section of the transportation systems and are therefore useful to identify the needs for achieving a frictionless passenger experience and improving the cyber-security toward a really integrated pan-European multi-modal transportation environment.

The data analytics components are privacy-aware pluggable services available in the analytics toolbox of the IAI subsystem of the E-CORRIDOR framework. For each component, the proposed approach and technology have been framed with respect to the state of the art and data format and platform requirements. The match of the analytics with the pilot requirements has been expressed through the use cases, from which it is possible to notice the active participation of the pilots in the component design. Moreover, the important role covered by the analytics in achieving the needs of the multi-modal transportation and the project objectives has been outlined. Finally, synergies among analytics and advanced security services have been presented with the purpose of better exploiting the capabilities of all the components available in the E-CORRIDOR framework.

The next period will be devoted to the refinement of the component features and requirements, and their development. Results on the first maturation of the data analytics techniques and their initial integration in the E-CORRIDOR framework will be reported in the next deliverable (D7.2, “Data Analytics techniques first maturation”) along with some preliminary demonstration.

11. References

[1] Oracle, "JAR File Overview," 2012. [Online]. Available: https://docs.oracle.com/javase/8/docs/technotes/guides/jar/jarGuide.html.

[2] Docker, Inc., "Docker: OS-level virtualization software," 2013. [Online]. Available: https://www.docker.com/.

[3] R. T. Fielding, "Architectural Styles and the Design of Network-based Software Architectures," in *Chapter 5: Representational State Transfer (REST)*, University of California, Irvine, 2000.

[4] OpenAPI Specification, "OpenAPI Initiative," 2020. [Online]. Available: https://spec.openapis.org/oas/v3.0.3.

[5] M. Fowler and R. Parsons, Domain Specific Languages, Addison-Wesley Professional, 2010.

[6] Yaml.org, "YAML: YAML Ain't Markup Language," [Online]. Available: https://yaml.org/. [Accessed 27 May 2021].

[7] HDFS, "Hadoop Distributed File System (HDFS)," 2006. [Online]. Available: https://hadoop.apache.org/docs/current/hadoop-project-dist/hadoop-hdfs/HdfsDesign.html.

[8] H. Mihaly, "From NASA to EU: the evolution of the TRL scale in Public Sector Innovation," *The Innovation Journal*, pp. 1-23, 2017.

[9] U. Fugiglando, P. Santi, S. Milardo, K. Abida and C. Ratti, "Characterizing the "driver dna" through can bus data analysis," in *Proceedings of the 2Nd ACM International Workshop on Smart, Autonomous, and Connected Vehicular Systems and Services, CarSys '17*, New York, NY, USA, 2017.

[10] P. Bonsall, R. Liu and W. Young, "Modelling safety-related driving behaviour—impact of parameter values," *Transportation Research Part A: Policy and Practice*, pp. 425-444, 2005.

[11] L. Eboli, G. Mazzulla and G. Pungillo, "ombining speed and acceleration to define car users' safe or unsafe driving behaviour," *Transportation Research Part C: Emerging Technologies*, pp. 113-125, 2016.

[12] A. S. Zeeman and M. J. Booysen, "Combining speed and acceleration to detect reckless driving in the informal public transport industry," in *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, The Hague, The Netherlands, 2013.

[13] D. Filev, J. Lu, F. Tseng and K. Prakah-Asante, "Real-time driver characterization during car following using stochastic evolving models," in *2011 IEEE International Conference on Systems, Man, and Cybernetics*, 2011.

[14] A. Rønning, "Rewarding safe drivers could make roads safer," 12 September 2013. [Online]. Available: https://sciencenordic.com/cars-and-traffic-forskningno-insurance/rewarding-safe-drivers-could-make-roads-safer/1390317.

[15] G. Costantino, F. Martinelli, P. Santi and I. Matteucci, "A Privacy-Preserving Infrastructure for Driver's Reputation Aware Automotive Services," in *Proceedings*

Workshop on Socio-Technical Aspects in Security (STAST), Luxembourg City, Luxembourg, 2019.

[16] R. Fracchia and M. Meo, "Analysis and design of warning delivery service in intervehicular networks," *IEEE Transactions on Mobile Computing*, pp. 832-845, 2008.

[17] G. Costantino, R. R. Maiti, F. Martinelli and P. Santi, "Private mobility-cast for opportunistic networks," *Computer Networks*, pp. 28-42, 2017.

[18] GSA, "GNSS Market Report," 2019. [Online]. Available: https://www.gsa.europa.eu/system/files/reports/market_report_issue_6_v2.pdf.

[19] H. S. Maghdid, I. A. Lami, K. Z. Ghafoor and J. Lloret, "Seamless outdoors-indoors localization solutions on smartphones: Implementation and challenges.," 2016.

[20] P. Casale, O. Pujol and P. Radeva, "Personalization and user verification in wearable systems using biometric walking patterns.," in *Personal and Ubiquitous Computing*, 2012.

[21] U. Mahbub, S. Sarkar, V. M. Patel and R. Chellappa, "Active user authentication for smartphones: A challenge data set and benchmark results.," in *IEEE 8th international conference on biometrics theory, applications and systems*, 2016.

[22] M. D. Papamichail, K. C. Chatzidimitriou, T. Karanikiotis, N. C. I. Oikonomou, A. L. Symeonidis and S. K. Saripalle, "BrainRun: A behavioral biometrics dataset towards continuous implicit authentication.," in *Data*, 2019.

[23] P. Baronti, P. Barsocchi, S. Chessa, F. Mavilia and F. Palumbo, "Indoor bluetooth low energy dataset for localization, tracking, occupancy, and social interaction.," 2018.

[24] A. Achroufene, Y. Amirat and A. Chibani, "RSS-based indoor localization using belief function theory.," in *IEEE Transactions on Automation Science and Engineering*, 2018.

[25] A. Achroufene, A. Chibani and Y. Amirat, "Using Dempster-Shafer Theory for RSS-based Indoor Localization.," in *IEEE International Conference on Fuzzy Systems*, 2020.

[26] Apple-Google, "Privacy-Preserving Contact Tracing," 2020. [Online]. Available: https://covid19.apple.com/contacttracing.

[27] Airport Technology, "Malpensa airport goes live with new Bluetooth passenger tracking system," 2012. [Online]. Available: https://www.airport-technology.com/uncategorised/newsmalpensa-airport-goes-live-new-bluetooth-passenger-tracking-system/.

[28] Google, "Eddystone protocol," 2018. [Online]. Available: https://github.com/google/eddystone.

[29] N. Sulman, T. Sanocki, D. Goldgof and R. Kasturi, "How effective is human video surveillance performance?," in *19th International Conference on Pattern Recognition*, Tampa, 2008.

[30] M. Y. Abbass, K.-C. Kwon, N. Kim, S. A. Abdelwahab, F. E. A. El-Samie and A. A. M. Khalaf, "Efficient object tracking using hierarchical convolutional features model and correlation filters," *The Visual Computer*, p. 831–842, 2021.

[31] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 961-971, 2016.

[32] W. Luo, J. Xing, A. Milan, X. Zhang, W. Liu and T.-K. Kim, "Multiple object tracking: A literature review," *Artificial Intelligence*, 2021.

[33] Y. Lin, L. Xie, Y. Wu, C. Yan and Q. Tian, "Unsupervised Person Re-identification via Softened Similarity Learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020.

[34] W. Li, R. Zhao, T. Xiao and X. Wang, "DeepReID: Deep Filter Pairing Neural Network for Person Re-identification," *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[35] W.-S. Zheng, S. Gong and T. Xiang, "Associating Groups of People," in *Conference: British Machine Vision Conference, BMVC*, 2009.

[36] S. Ardeshir and A. Borji, "Ego2Top: Matching Viewers in Egocentric and Top-View Videos," in *European Conference on Computer Vision (ECCV)*, 2016.

[37] A. Bochkovskiy, C.-Y. Wang and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv e-prints*, 2020.

[38] N. Wojke, A. Bewley and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," *IEEE International Conference on Image Processing (ICIP)*, 2017.

[39] A. Buriro, B. Crispo, F. Delfrari and K. Wrona, "Hold and sign: a novel behavioral biometrics for smartphone user authentication," in *2016 IEEE Security and Privacy Workshops (SPW)*, 2016.

[40] P. Fernandez-Lopez, J. Liu-Jimenez, C. Sanchez-Redondo and R. Sanchez-Reillo, "Gait recognition using smartphone," in *2016 IEEE International Carnahan Conference on Security Technology (ICCST)*, 2016.

[41] M. N. Malik, M. A. Azam, M. Ehatisham-Ul-Haq, W. Ejaz and A. Khalid, "ADLAuth: Passive Authentication Based on Activity of Daily Living Using Heterogeneous Sensing in Smart Cities," *Sensors*, 2019.

[42] M. Abuhamad, T. Abuhmed, D. Mohaisen and D. Nyang, "AUToSen: Deep-Learning-Based Implicit Continuous Authentication Using Smartphone Sensors," *IEEE Internet of Things Journal*, pp. 5008-5020, 2020.

[43] Q. Zou, Y. Wang, Q. Wang, Y. Zhao and Q. Li, "Deep Learning-Based Gait Recognition Using Smartphones in the Wild," *IEEE Transactions on Information Forensics and Security*, pp. 3197-3212, 2020.

[44] Z. Sitová, J. Šeděnka, Q. Yang, G. Peng, G. Zhou, P. Gasti and K. S. Balagani, "HMOG: New Behavioral Biometric Features for Continuous Authentication of Smartphone Users," *IEEE Transactions on Information Forensics and Security*, pp. 877-892, 2016.

[45] G. Giorgi, A. Saracino and F. Martinelli, "Using Recurrent Neural Networks for Continuous Authentication through Gait Analysis," *Pattern Recognition Letters*, 2021[InPress].

[46] Y. Kortli, M. Jridi, A. Falou and M. Atri, "Face Recognition Systems: A Survey," *Sensors,* 2020.

[47] H. Yang and X. Wang, "Cascade classifier for face detection," *Journal of Algorithms Computational Technology,* p. 187–197, 2016.

[48] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition,* Kauai, HI, 2001.

[49] Y. Ouerhani, A. Alfalou and C. Brosseau, "Road mark recognition using HOG-SVM and correlation," in *Optics and Photonics for Information Processing XI,* San Diego, California, USA, 2017.

[50] J. Rettkowski, A. Boutros and D. Göhringer, "HW/SW Co-Design of the HOG algorithm on a Xilinx Zynq SoC," *Journal of Parallel and Distributed Computing,* pp. 50-62, 2017.

[51] H. Seo and P. Milanfar, "Face verification using the lark representation," *IEEE Transactions on Information Forensics and Security,* pp. 1275-1286, 2011.

[52] J. Shah, M. Sharif, M. Raza and A. Azeem, "A Survey: Linear and Nonlinear PCA Based Face Recognition Techniques," *International Arab Journal of Information Technology,* p. 536–545, 2013.

[53] G. Du, F. Su and A. Cai, "Face recognition using SURF features," in *MIPPR 2009: Pattern Recognition and Computer Vision; International Society for Optics and Photonics,* Bellingham, WA, USA, 2009.

[54] M. Calonder, V. Lepetit, M. Ozuysal, T. Trzcinski, C. Strecha and P. Fua, "BRIEF: Computing a Local Binary Descriptor Very Fast," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1281-1298, 2012.

[55] Q. Wang, D. Xiong, A. Alfalou and C. Brosseau, "Optical image authentication scheme using dual polarization decoding configuration," *Optics and Lasers in Engineering,* p. 151–161, 2019.

[56] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience,* pp. 71-86, 1991.

[57] M. Annalakshmi, S. Roomi and A. Naveedh, "A hybrid technique for gender classification with SLBP and HOG features," *Cluster Computing,* pp. 11-20, 2019.

[58] A. Vinay, D. Hebbar, V. Shekhar, K. Murthy and S. Natarajan, "Two novel detector-descriptor based approaches for face recognition using sift and surf," *Procedia Computer Science,* pp. 185-197, 2015.

[59] S. Hussain, T. Napoléon and F. Jurie, "Face Recognition Using Local Quantized Patterns," in *British Machine Vision Conference,* 2012.

[60] F. Smach, J. Miteran, M. Atri, J. Dubois, M. Abid and J. Gauthier, "An FPGA-based accelerator for Fourier Descriptors computing for color object recognition using SVM," *Journal of Real-Time Image Processing,* p. 249–258, 2007.

[61] T. Napoléon and A. Alfalou, "Pose invariant face recognition: 3D model from single photo," *Optics and Lases in Engineering,* p. 150–161, 2017.

[62] A. HajiRassouliha, T. Gamage, M. Parker, M. Nash, A. Taberner and P. Nielsen, "FPGA implementation of 2D cross-correlation for real-time 3D tracking of deformable surfaces," in *2013 28th International Conference on Image and Vision Computing New Zealand (IVCNZ 2013)*, Piscataway, NJ, USA, 2013.

[63] Y. Ouerhani, M. Jridi, A. Alfalou and C. Brosseau, "Optimized pre-processing input plane GPU implementation of an optical face recognition technique using a segmented phase only composite filter," *Optics Communications*, pp. 33-44, 2013.

[64] A. Alfalou and C. Brosseau, "Understanding Correlation Techniques for Face Recognition: From Basics to Applications," in *Face Recognition*, Rijeka, Croatia, IntechOpen, 2010.

[65] T. Napoléon and A. Alfalou, "Local binary patterns preprocessing for face identification/verification using the VanderLugt correlator," in *Optical Pattern Recognition XXV; International Society for Optics and Photonics*, Bellingham, WA, USA, 2014.

[66] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," in *IEEE conference on computer vision and pattern recognition*, Boston, MA, USA, 2015.

[67] I. Kambi Beli and C. Guo, "Enhancing face identification using local binary patterns and k-nearest neighbors," *Journal of Imaging*, 2017.

[68] P. Khoi, L. Thien and V. Viet, "Face Retrieval Based on Local Binary Pattern and Its Variants: A Comprehensive Study," *International Journal of Advanced Computer Science and Applications*, p. 249–258, 2016.

[69] M. Karaaba, O. Surinta, L. Schomaker and M. Wiering, "Robust face recognition by computing distances from multiple histograms of oriented gradients," in *IEEE Symposium Series on Computational Intelligence*, Piscataway, NJ, USA, 2015.

[70] A. Alahi, R. Ortiz and P. Vandergheynst, "Freak: Fast retina keypoint," in *IEEE Conference on Computer Vision and Pattern Recognition*, Providence, RI, USA, 2012.

[71] A. Lima, H. Zen, Y. Nankaku, C. Miyajima, K. Tokuda and T. Kitamura, "On the use of kernel PCA for feature extraction in speech recognition," in *8th European Conference on Speech Communication and Technology, EUROSPEECH 2003 - INTERSPEECH 2003*, Geneva, Switzerland.

[72] K. Simonyan, O. Parkhi, A. Vedaldi and A. Zisserman, "Fisher vector faces in the wild," in *BMVC 2013—British Machine Vision Conference*, Bristol, UK, 2013.

[73] V. Perlibakas, "Face recognition using principal component analysis and log-gabor filters," arXiv, arXiv:cs/0605025, 2006.

[74] Z. Huang, W. Li, J. Shang, J. Wang and T. Zhang, "Non-uniform patch based face recognition via 2D-DWT," *Image and Vision Computing*, pp. 12-19, 2015.

[75] Z. Sufyanu, F. Mohamad, A. Yusuf and M. Mamat, "Enhanced Face Recognition Using Discrete Cosine Transform," *Engineering Letters*, p. 52–61, 2016.

[76] A. Vinay, V. Shekhar, K. Murthy and S. Natarajan, "Performance study of LDA and KFA for gabor based face recognition system," *Procedia Computer Science*, p. 960–969, 2015.

[77] S. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," *IEEE Transactions on Information Forensics and Security,* pp. 2100-2109, 2014.

[78] A. Fathima, S. Ajitha, V. Vaidehi, M. Hemalatha, R. Karthigaiveni and R. Kumar, "Hybrid approach for face recognition combining GaborWavelet and Linear Discriminant Analysis," in *IEEE International Conference on Computer Graphics, Vision and Information Security (CGVIS)*, Bhubaneswar, India, 2015.

[79] O. Barkan, J. Weill, L. Wolf and H. Aronowitz, "Fast high dimensional vector multiplication face recognition," in *IEEE International Conference on Computer Vision*, Sydney, Australia, 2013.

[80] F. Juefei-Xu, K. Luu and M. Savvides, "Spartans: Single-sample periocular-based alignment-robust recognition technique applied to non-frontal scenarios," *IEEE Transactions on Image Processing,* pp. 4780-4795, 2015.

[81] Y. Yan, H. Wang and D. Suter, "Multi-subregion based correlation filter bank for robust face recognition," *Pattern Recognition*, p. 3487–3501, 2014.

[82] C. Ding and D. Tao, "Robust face recognition via multimodal deep face representation," *IEEE Transactions on Multimedia*, p. 2049–2058, 2015.

[83] R. Sharma and M. Patterh, "A new pose invariant face recognition system using PCA and ANFIS," *Optik,* pp. 3483-3487, 2015.

[84] M. Moussa, M. Hmila and A. Douik, "A Novel Face Recognition Approach Based on Genetic Algorithm Optimization," *Studies in Informatics and Control*, p. 127–134, 2018.

[85] A. Mian, M. Bennamoun and R. Owens, "An efficient multimodal 2D-3D hybrid approach to automatic face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, p. 1927–1943, 2007.

[86] H. Cho, R. Roberts, B. Jung, O. Choi and S. Moon, "An efficient hybrid face recognition algorithm using PCA and GABOR wavelets," *International Journal of Advanced Robotic Systems,* 2014.

[87] D. Guru, M. Suraj and S. Manjunath, "Fusion of covariance matrices of PCA and FLD," *Pattern Recognition Letter ,* p. 432–440, 2011.

[88] J. Sing, S. Chowdhury, D. Basu and M. Nasipuri, "An improved hybrid approach to face recognition by fusing local and global discriminant features," *International Journal of Biometrics ,* pp. 144-164, 2012.

[89] P. Kamencay, M. Zachariasova, R. Hudec, R. Jarina, M. Benco and J. Hlubik, "A novel approach to face recognition using image segmentation based on spca-knn method," *Radioengineering*, pp. 92-99, 2013.

[90] J. Sun, Y. Fu, S. Li, J. He, C. Xu and L. Tan, "Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors," *Journal of Sensors*, 2018.

[91] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[92] L. M. Dang, K. Min, H. Wang, P. M. J. C. H. Lee and H. 108Moon, "Sensor-based and vision-based human activity recognition: A comprehensive survey," *Pattern Recognition,* vol. 108, 2020.

[93] A. Bux, P. Angelov and Z. Habib, "Vision based human activity recognition: a review," *Advances in Computational Intelligence Systems,* pp. 341-371, 2017.

[94] J. K. Aggarwal and L. Xia, "Human activity recognition from 3d data: A review," *Pattern Recognition Letters,* pp. 70-80, 2014.

[95] F. Han, B. Reily, W. Hoff and H. Zhang, "Space-time representation of people based on 3D skeletal data: A review," *Computer Vision and Image Understanding,* pp. 85-105, 2017.

[96] L. L. Presti and M. La Cascia, "3D skeleton-based human action classification: A survey," *Pattern Recognition,* pp. 130-147, 2016.

[97] B. Zhan, D. N. Monekosso, P. Remagnino, S. A. Velastin and L. Q. Xu, "Crowd analysis: a survey," *Machine Vision and Applications,* pp. 345-357, 2008.

[98] Y. Du, W. Wang and L. Wang, "Hierarchical recurrent neural network for skeleton based action recognition," in *IEEE conference on computer vision and pattern recognition*, 2015.

[99] W. Zhu, C. Lan, J. Xing, W. Zeng, Y. Li, L. Shen and X. Xie, "Co-occurrence feature learning for skeleton based action recognition using regularized deep LSTM networks," in *AAAI Conference on Artificial Intelligence*, 2016.

[100] "Routing/online routers," OpenStreetMap Wiki, 2021. [Online]. Available: https://wiki.openstreetmap.org/wiki/Routing/online_routers.

[101] "Google Transit basics-About Google Transit," 2021. [Online]. Available: https://support.google.com/transitpartners/answer/1111471?hl=en.

[102] "TripGo Web Trip Planner," 2021. [Online]. Available: https://tripgo.com/.

[103] “OpenTripPlanner 2,” 2020. [Online]. Available: http://docs.opentripplanner.org/en/latest/.

[104] "OSRM," 2021. [Online]. Available: http://project-osrm.org/.

[105] "openrouteservice - Services," 2020. [Online]. Available: https://openrouteservice.org/services/.

[106] "The GraphHopper Directions API Route Planning For Your Application," 2020. [Online]. Available: https://www.graphhopper.com/.

[107] "OpenStreetMap," OpenStreetMap, [Online]. Available: https://www.openstreetmap.org/about.

[108] "OTP Sandbox Extensions," [Online]. Available: http://docs.opentripplanner.org/en/latest/SandboxExtension/.

[109] “GTFS: Making Public Transit Data Universally Accessible,” [Online]. Available: https://gtfs.org/.

[110] “GTFS Realtime Reference v2,” [Online]. Available: https://gtfs.org/reference/realtime/v2/.

[111] M. v. d. Tuin, T. Hubers, P. Tekieli and A. Croockewit, "OpenTripPlanner: A multimodal tripplanner," 2016. [Online]. Available: https://delftswa.gitbooks.io/desosa2016/content/opentripplanner/chapter.html.

[112] P. E. Hart, N. J. Nilsson and B. Raphael, "A Formal Basis for the Heuristic Determination of Minimum Cost Paths," *IEEE Transactions on Systems Science and Cybernetics* , pp. 100-107, 1968.

[113] D. Delling, T. Pajor and R. F. Werneck, "Round-based public transit routing," *Transportation Science*, pp. 591-604, 2015.

[114] "General Transit Feed Specification Reference," 2019. [Online]. Available: https://gtfs.org/reference/static.

[115] "GTFS-Flex routing," [Online]. Available: http://docs.opentripplanner.org/en/v1.5.0/Flex/.

[116] "General Bikeshare Feed Specification," 2021. [Online]. Available: https://github.com/NABSA/gbfs.

[117] Z. Brakerski, C. Gentry and V. Vaikuntanathan, "(Leveled) fully homomorphic encryption without bootstrapping," in *ITCS '12: Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, Cambridge, Massachusetts, 2012.

[118] Z. Brakerski, "Fully homomorphic encryption without modulus switching from classical gapsvp," in *CRYPTO 2012: Advances in Cryptology*, 2012, 2012.

[119] J. Fan and F. Vercauteren, "Somewhat Practical Fully Homomorphic Encryption," Cryptology ePrint Archive, Report 2012/144, 2012.

[120] CEA-LIST, "Cingulata – Homomorphic encryption technology," 2018. [Online]. Available: https://github.com/CEA-LIST/Cingulata.

[121] I. Chillotti, N. Gama, M. Georgieva and M. Izabachène, "Faster fully homomorphic encryption: Bootstrapping in less than 0.1 seconds," in *Advances in Cryptology -- ASIACRYPT 2016*, 2016.

[122] I. Chillotti, N. Gama, M. Georgieva and M. Izabachène, "Faster Packed Homomorphic Operations and Efficient Circuit Bootstrapping for TFHE," in *ASIACRYPT 2017: Advances in Cryptology*, 2017.

[123] C. Aguilar-Melchor, S. Fau, C. Fontaine, G. Gogniat and R. Sirdey, "Recent Advances in Homomorphic Encryption: A Possible Future for Signal Processing in the Encrypted Domain," *IEEE Signal Processing Magazine,* pp. 108-117, 2013.

[124] S. Fau, R. Sirdey, C. Fontaine, C. Melchor and G. Gogniat, "Towards Practical Program Execution over Fully Homomorphic Encryption Schemes," in *2013 Eighth International Conference on P2P, Parallel, Grid, Cloud and Internet Computing*, 2013.

[125] D. Micale, G. Costantino, I. Matteucci, G. Patanè and G. Bella, "Secure Routine: A Routine-Based Algorithm for Drivers Identification," in *The Ninth International Conference on Advances in Vehicular Systems, Technologies and Applications*, 2020.

[126] C. A. da Silveira Barreto, "OBDdatasets," 2018. [Online]. Available: https://github.com/cephasax/OBDdatasets/.

[127] HCRL Hacking and Countermeasure Research Lab, "Driving Dataset," [Online]. Available: http://ocslab.hksecurity.net/Datasets/.

[128] F. Martinelli, F. Mercaldo, V. Nardone, A. Orlando and A. Santone, "Who's Driving My Car? A Machine Learning based Approach to Driver Identification," in *4th International Conference on Information Systems Security and Privacy*, 2018.

[129] F. Martinelli, F. Mercaldo, A. Orlando, V. Nardone, A. Santone and A. K. Sangaiah, "Human behavior characterization for driving style recognition in vehicle system," *Computers & Electrical Engineering,* 2020.

[130] K. Uvarov and A. Ponomarev, "Driver Identification with OBD-II PublicData," in *28th Conference of Open Innovations Association (FRUCT)*, 2021.

[131] J. Feng, C. Rong, F. Sun, D. Guo and Y. Li, "PMF: A Privacy-preserving Human Mobility Prediction Framework via Federated Learning," in *ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 2020.

[132] H. B. McMahan, E. Moore, D. Ramage, S. Hampson and B. Agüera y Arcas, "Communication-Efficient Learning of Deep Networks from Decentralized Data," arXiv, 2016.

[133] E. Boyle, N. Gilboa and Y. Ishai, "Function Secret Sharing: Improvements and Extensions," in *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016.

[134] O. Goldreich, "Secure Multi-Party Computation. Manuscript. Preliminary," Department of Computer Science and Applied Mathematics, Weizmann Institute of Science, Rehovot, Israel, 1999.

[135] T. Ryffel, D. Pointcheval and F. Bach, "ARIANN: Low-Interaction Privacy-Preserving Deep Learning via Function Secret Sharing," arXiv, 2020.

[136] W. Stallings, Network Security Essentials: Applications and Standards, Pearson, 2016.

[137] "A European Strategy for low-emission mobility," 2016. [Online]. Available: https://ec.europa.eu/clima/policies/transport_en.

[138] J. Mulrow, K. Machaj, J. Deanes and S. Derrible, "The state of carbon footprint calculators: An evaluation of calculator design and user interaction features," *Sustainable Production and Consumption*, vol. 18, pp. 33-40, 2019.

[139] "GREEN DRIVING TOOL," [Online]. Available: https://green-driving.jrc.ec.europa.eu/.

[140] S. Nie, L. Liu and Y. Du, "Free-fall: hacking tesla from wireless to CAN bus," *Briefing, Black Hat USA,* 2017.

[141] M. Müter, A. Groll and F. C. Freiling, "A structured approach to anomaly detection for in-vehicle networks," in *Sixth International Conference on Information Assurance and Security, {IAS} 2010*, Atlanta, GA, USA, 2010.

[142] U. E. Larson, D. Nilsson and E. Jonsson, "An approach to specification-based attack detection for in-vehicle networks," in *Intelligent Vehicles Symposium, 2008 IEEE*, 2008.

[143] T. Hoppe, S. Kiltz and J. Dittmann, "Security threats to automotive CAN networks -- Practical examples and selected short-term countermeasures," in *Reliability Engineering \& System Safety*, 2011.

[144] I. Studnia, E. Alata, V. Nicomette, M. Kaâniche and Y. Laarouchi, "A language-based intrusion detection approach for automotive embedded networks," in *The 21st IEEE Pacific Rim International Symposium on Dependable Computing (PRDC 2015)*, 2014.

[145] S. N. Narayanan, S. Mittal and A. Joshi, "Using semantic technologies to mine vehicular context for security," in *2016 IEEE 37th Sarnoff Symposium*, 2016.

[146] T. Islinger, Y. Mori, J. Neumüller, M. Prisching and R. Schmidt, "Autosar SecOC for CAN FD," *CAN Newsletter*, 2017.

[147] L. W. Choi, K. Joo, H. J. Jo, M. C. Park and D. H. Lee, "VoltageIDS: Low-Level Communication Characteristics for Automotive Intrusion Detection System," *IEEE Transactions on Information Forensics and Security*, 2018.

[148] K.-T. Cho and K. G. Shin, "Fingerprinting Electronic Control Units for Vehicle Intrusion Detection," in *25th USENIX Security Symposium, USENIX Security 16*, 2016.

[149] A. Taylor, N. Japkowicz and S. Leblanc, "Frequency-based anomaly detection for the automotive CAN bus," in *2015 World Congress on Industrial Control Systems Security (WCICSS)*, 2015.

[150] H. M. Song, H. R. Kim and H. K. Kim, "Intrusion detection system based on the analysis of time intervals of CAN messages for in-vehicle network," in *International Conference on Information Networking*, 2016.

[151] Z. Wei, Y. Yang, Y. Rehana, Y. Wu, J. Weng and R. H. Deng, "IoVShield: An Efficient Vehicular Intrusion Detection System for Self-driving," in *Information Security Practice and Experience: 13th International Conference, ISPEC 2017*, 2017.

[152] A. Theissler, "Anomaly detection in recordings from in-vehicle networks," *Big data and applications*, 2014.

[153] M.-J. Kang and J.-W. Kang, "Intrusion detection system using deep neural network for in-vehicle network security," *PloS one*, 2016.

[154] M. Levi, Y. Allouche and A. Kontorovich, "Advanced Analytics for Connected Cars Cyber Security," *CoRR*, 2017.

[155] R. Rieke, M. Seidemann, E. K. Talla, D. Zelle and B. Seeger, "Behavior Analysis for Safety and Security in Automotive Systems," in *Parallel, Distributed and Network-Based Processing (PDP), 2017 25nd Euromicro International Conference on*, 2017.

[156] N. Nowdehi, W. Aoudi, M. Almgren and T. Olovsson, "CASAD: CAN-Aware Stealthy-Attack Detection for In-Vehicle Networks," arXiv, 2019.

[157] D. Stabili, M. Marchetti and M. Colajanni, "Detecting attacks to internal vehicle networks through hamming distance," in *2017 AEIT International Annual Conference*, 2017.

[158] M. Marchetti and D. Stabili, "Anomaly detection of CAN bus messages through analysis of ID sequences," in *IEEE Intelligent Vehicles Symposium, IV 2017*, Los Angeles, CA, USA, 2017.

[159] Y. Chevalier, R. Rieke, F. Fenzl, A. Chechulin and I. Kotenko, "ECU-Secure: Characteristic Functions for In-Vehicle Intrusion Detection," in *International Symposium on Intelligent and Distributed Computing*, 2019.

[160] A. Taylor, S. P. Leblanc and N. Japkowicz, "Probing the Limits of Anomaly Detectors for Automobiles with a Cyber Attack Framework," *IEEE Intelligent Systems,* 2018.

[161] I. Berger, R. Rieke, M. Kolomeets, A. Chechulin and I. Kotenko, "Comparative Study of Machine Learning Methods for In-Vehicle Intrusion Detection," in *Computer Security. ESORICS 2018 International Workshops, CyberICPS 2018 and SECPRE 2018*, 2019.

[162] O. Y. Al-Jarrah, C. Maple, M. Dianati, D. Oxtoby and A. Mouzakitis, "Intrusion Detection Systems for Intra-Vehicle Networks: A Review," *IEEE Access,* 2019.

[163] G. Dupont, J. den Hartog, S. Etalle and A. Lekidis, "A Survey of Network Intrusion Detection Systems for Controller Area Network," in *2019 IEEE International Conference of Vehicular Electronics and Safety (ICVES)*, 2019.

[164] F. Fenzl, R. Rieke, Y. Chevalier, A. Dominik and I. Kotenko, "Continuous Fields: Enhanced In-Vehicle Anomaly Detection using Machine Learning Models," *Simulation Modelling Practice and Theory,* 2020.

[165] Hacking and Countermeasure Research Lab (HCRL), "Car-Hacking Dataset for the intrusion detection," http://ocslab.hksecurity.net/Datasets/CAN-intrusion-dataset, 2018.

[166] M. E. Verma, M. D. Iannacone, R. A. Bridges, S. C. Hollifield, B. Kay and F. L. Combs, "ROAD: The Real ORNL Automotive Dynamometer Controller Area Network Intrusion Detection Dataset (with a comprehensive CAN IDS dataset survey & guide)," eprint arXiv:2012.14600, 2020.

[167] C. Miller and C. Valasek, "A survey of remote automotive attack surfaces," Black Hat USA, 2014.

[168] Tencent Keen Security Lab, "New Vehicle Security Research by KeenLab: Experimental Security Assessment of BMW Cars," 22 May 2018. [Online]. Available: https://keenlab.tencent.com/en/2018/05/22/New-CarHacking-Research-by-KeenLab-Experimental-Security-Assessment-of-BMW-Cars/.

[169] Tencent Keen Security Lab, "Tencent Keen Security Lab: Experimental Security Assessment on Lexus Cars," 30 March 2020. [Online]. Available: https://keenlab.tencent.com/en/2020/03/30/Tencent-Keen-Security-Lab-Experimental-Security-Assessment-on-Lexus-Cars/.

[170] G. Bella, P. Biondi, G. Costantino and I. Matteucci, "Are you secure in your car?: poster," in *Proceedings of the 12th Conference on Security and Privacy in Wireless and Mobile Networks*, Miami, Florida, USA, 2019.

[171] S.-F. Lokman, A. T. Othman and M.-H. Abu-Bakar, "Intrusion detection system for automotive controller area network (can) bus system: a review," *EURASIP Journal on Wireless Communications and Networking,* 2019.

[172] T. Hoppe, S. Kiltz and J. Dittmann, "Applying intrusion detection to automotive IT - early insights and remaining challenges," *Journal of Information Assurance and Security (JIAS),* pp. 226-235, 2009.

[173] R. W. Hamming, "Error detecting and error correcting codes," *The Bell System Technical Journal,* pp. 147-160, 1950.

[174] The OBDII Home Page, "OBD-II Background," 2011. [Online]. Available: http://www.obdii.com/background.html.

[175] D. Clegg and R. Barker, Case Method Fast-Track: A RAD Approach, Addiso-Wesley, 1994.

[176] "GTFS Realtime Reference v2," [Online]. Available: https://gtfs.org/reference/realtime/v2/.

A. Appendix

A.1 Definitions and Abbreviations

Term	Meaning
AAL	Ambient Assisted Living
ADAS	Advanced Driver-Assistance Systems
AI	Artificial Intelligence
API	Application Programming Interface
ASI	Advanced Security Services Infrastructure (E-CORRIDOR framework)
AT	Airport-Train (E-CORRIDOR pilot)
AUTOSAR	Automotive Open System Architecture
BLE	Bluetooth Low Energy
BYOD	Bring Your Own Device
CAN bus	Controller Area Network
CCTV	Closed-circuit television
CF	Correlation filters
CGW	Central Gateway
CO2	Carbon dioxide – air pollutant
CNN	Convolutional Neural Networks
CSV	Comma-separated values
CTI	Cyber-Threat Information
CVE	Common Vulnerabilities and Exposures
CVSS	Common Vulnerability Scoring System
DBC	Database CAN file
DCT	Discrete Cosine Transform
DLI	DSA Lifecycle Infrastructure (E-CORRIDOR framework)
DMO	Data Manipulation Operation
DNA	Here, as an analogy to the molecule that carries the genetic instructions
DoA	Description of the Action
DPO	Data Protected Object
DPOS	Data Protected Object Storage
DSA	Data Sharing Agreement
DTW	Dynamic Time Warping
DWT	Discrete Wavelet Transform

ECU	Electronic Control Unit
EEG	Electro-encephalography
ETSI	European Telecommunications Standards Institute
EU	European Union
FHE	Fully Homomorphic Encryption
FL	Federated Learning
e-wallet	Digital wallet
GHG	Greenhouse gas
GNSS	Global Navigation Satellite Systems
GPS	Global Positioning System
GTFS	General Transit Feed Specification
GRU	Gated recurrent units
HAR	Human Action Recognition
HCRL	Hacking and Countermeasure Research lab
HDFS	Hadoop Distributed File System
HMOG	Hand Movement, Orientation, and Grasp
HOG	Histogram of Oriented Gradient
IAI	Information Analytics Infrastructure (E-CORRIDOR framework)
ICA	Independent component analysis
ICT	Information and Communications Technology
IDS	Intrusion Detection System
IoT	Internet of Things
IP	Internet Protocol
IPS	Intrusion Prevention System
IR	Infrared
IMU	Inertial measurement unit
ISAC	Information Sharing and Analytics Center
ISI	Information Sharing Infrastructure (E-CORRIDOR framework)
ITS	Intelligent Transportation System
IVI	In-Vehicle Infotainment (IVI)
JAR	Executable Java
JRC	Joint Research Centre
JSON	JavaScript Object Notation
LBP	Local binary pattern

LDA	Linear Discriminant Analysis
LPQ	Local phase quantization
LSTM	Long Short-Term Memory
MEMS	Micro Electro-Mechanical Systems
ML	Machine Learning
MoSCoW	Must have, Should have, Could have, and Won't have but would like
MMT-ISAC	Information Sharing and Analysis Center for Multi-Modal transport (E-CORRIDOR pilot)
MPC	Secure Multi-Party Computation
NED	US National Elevation Dataset
NN	Neural Network
OBD	On-board diagnostics
OCSVM	One-class support vector machine
OEM	Original Equipment Manufacturer
ORNL	Oak Ridge National Laboratory
ORS	Open Route Service
OSM	Open Street Map
OSRM	Open Source Routing Machine
OTP	Open Trip Planner
PCA	Principal Component Analysis
PRM	People with Reduced Mobility
PSR	Private Secure Routine
R&D	Research and Development
REST	Representational state transfer
RFID	Radio-frequency identification
RGB	Red, green and blue – color model
RGB-D	Red, green, blue and depth
RNN	Recurrent Neural Network
RPM	Revolutions per minute
RSSI	Received Signal Strength Indicator
SD	Secure Digital – non-volatile memory card
SDK	Software development kit
SecOC	Secure Onboard Communication
SIEM	Security Information And Event Management

SIFT	Scale-invariant feature transform
SSO	Single Sign-On
SSR	Special Service Request
STIX	Structured threat information expression
S2C	Smart cities and car sharing (E-CORRIDOR pilot)
SVM	Support Vector Machine
TRL	Technology Readiness Level
2PC	Secure Two Party Computation
UC	Use case
UDS	Unified Diagnostic Services
US	User story
V2I	Vehicle to Infrastructure
V2V	Vehicle to Vehicle
XML	Extensible Markup Language

A.2 Details on specific Data Formats

A.2.1 Examples of CAN bus Datasets

In the following we show excerpts of different public CAN data sets and how the required information is logged there.

```
(1000000005.421648) can0 230#FF000002EC000000
(1000000005.423523) can0 2E1#0000000000000004
(1000000005.424562) can0 354#2009C00000026D80
(1000000005.424563) can0 5E1#893FE0080A000C80
(1000000005.424564) can0 00E#2052960208097564
(1000000005.424566) can0 25B#FFFFFFFFFFFFFFFF
(1000000005.425584) can0 162#00080003EA11F4CE
(1000000005.425585) can0 69E#044004811FC01542
(1000000005.425586) can0 125#9000415F433E8160
(1000000005.425588) can0 0A7#1090FCA4D12DD0A0
```

Figure 20 Automotive Dynamometer (ROAD) CAN Intrusion dataset (ORNL)

```
(1508687283.896772) slcan0 211#4E3718D7180000
(1508687283.896831) slcan0 45C#00000000
(1508687283.897208) slcan0 214#FBFE
(1508687283.897218) slcan0 5DF#0084
(1508687283.900328) slcan0 12E#C580027FD0FFFF00
(1508687283.900357) slcan0 090#1A000000
(1508687283.901062) slcan0 0C6#751280148008BCA0
(1508687283.901072) slcan0 29A#00000000000007F8
(1508687283.901089) slcan0 2B7#00E0FFFE01
(1508687283.905236) slcan0 1F6#1E2082368005FFFE
```

Figure 21 Automotive CAN bus intrusion v2 (TU Eindhoven)

Figure 20 and Figure 21 show excerpts from datasets in *can-utils candump* format. Here all values except for the timestamp are recorded as hexadecimal values. The message ID is delimited from the message payload with a '#' symbol. The length of the payload is dependent on the vehicle and method recording. The ORNL dataset shown in Figure 20 always utilizes the maximum length of the CAN bus payload, whereas the CAN bus intrusion dataset from TU Eindhoven contains payload of variable length [166]. The *candump* format contains information on the bus where messages were recorded. There are no indicators on which message is an introduced intrusion, therefore an additional file containing metadata is required.

```
1  1478191642.238456,0131,8,db,7f,00,00,36,7f,0f,4e,R
2  1478191642.238700,0140,8,00,00,00,00,12,25,2f,39,R
3  1478191642.239006,0316,8,05,21,70,09,21,20,00,73,R
4  1478191642.239241,0316,8,45,29,24,ff,29,24,00,ff,T
5  1478191642.239473,0329,8,87,bd,7e,14,11,20,00,14,R
6  1478191642.239707,0316,8,45,29,24,ff,29,24,00,ff,T
7  1478191642.239947,0545,8,d8,00,00,8d,00,00,00,00,R
8  1478191642.240182,0316,8,45,29,24,ff,29,24,00,ff,T
9  1478191642.240506,0316,8,45,29,24,ff,29,24,00,ff,T
10 1478191642.240832,0316,8,45,29,24,ff,29,24,00,ff,T
```

Figure 22 HCRL Car-Hacking dataset

```
1  0.021265,1272,8,136,16,0,0,0,0,0,0,1
2  0.021507,504,8,246,4,255,239,254,0,12,13,1
3  0.023507,1503,1,168,108,100,109,112,108,117,118,1
4  0.025063,378,8,255,255,255,170,0,240,49,163,1
5  0.025309,382,8,255,255,255,0,255,64,0,255,1
6  0.025444,1420,7,0,0,0,0,0,127,224,118,1
7  0.025539,390,7,0,0,50,3,32,0,32,118,1
8  0.025758,1421,8,0,0,0,0,0,0,0,0,1
9  0.026052,1400,8,0,2,101,176,0,187,187,187,1
10 0.028393,302,8,199,127,255,127,224,255,255,0,1
```

Figure 23 Renault ZOE CAN bus dataset (Fraunhofer SIT)

Figure 22 and Figure 23 show excerpts from datasets in *csv* format. Here all individual values are delimited by comma. The open HCRL Car-Hacking dataset [165] provides log files with decimal ID values and a payload separated into byte-sized hexadecimal fields, whereas said fields in the Renault ZOE CAN bus dataset, provided by Fraunhofer SIT, are decimal. With this data format the maximum length of the payload is always fully used, but an additional value depicting the actual length of the payload field is provided.

All messages are also flagged whether they are valid messages sent from an ECU within the vehicle or originated from an intrusion scenario. Either directly introduced into the vehicle, as done in the HCRL dataset or synthetically introduced after recording, as done in the Renault ZOE dataset.

A.2.2 Examples of Alert Indicators in STIX Data Format

In the following we show some examples how alert information can be represented in STIX format.

```
{
    "created": "2020-01-26T17:55:10.442Z",
    "id": "identity--5206ba14-478f-4b0b-9a48-395f690c20a2",
    "identity_class": "system",
    "modified": "2020-01-26T17:55:10.442Z",
    "name": "Network IDS",
    "roles": [ "Cyber Security" ],
    "sectors": [ "technology" ],
    "spec_version": "2.1",
    "type": "identity"
}
```


Figure 24 STIX Format: Identity Example

The STIX Intrusion reporting format utilizes a model of the system, where every relevant component is provided with a specific unique *identity*. This *identity* contains meta information on roles and sectors of the component, such as shown in Figure 24.


```
{
    "count": 1,
    "created": "2020-01-26T17:55:10.442Z",
    "created_by_ref": "identity--5206ba14-478f-4b0b-9a48-395f690c20a2",
    "id": "sighting--8356e820-8080-4692-aa91-ecbe94006833",
    "modified": "2020-01-26T17:55:10.442Z",
    "observed_data_refs": [
        "observed-data--5046dbb6-0e6a-44bd-8b23-34ce41fae19e"
    ],
    "sighting_of_ref": "indicator--9299f726-ce06-492e-8472-2b52ccb53191",
    "spec_version": "2.1",
    "type": "sighting",
    "where_sighted_refs": [
        "identity--5206ba14-478f-4b0b-9a48-395f690c20a2"
    ]
}
```


Figure 25 STIX Format: Sighting Example

On the occurrence of a security relevant event a *sighting* event is created and assigned to a specific *identity*. An example for such an event is shown in Figure 25.

The object for such an event provides meta information, such as the relevant timestamps, number of previous occurrences of the same type, called *observed-data* (e.g., Figure 26), references to additionally provided data objects and a reference to a concrete *indicator* for the type of occurred incident.

```
{
    "id": "observed-data--5046dbb6-0e6a-44bd-8b23-34ce41fae19e",
    "object_refs": "file--f1b9f760-83d4-47ca-872b-7e46f2a9f54a",
...}
{
    "id": "file--f1b9f760-83d4-47ca-872b-7e46f2a9f54a",
    "content_ref": "artifact--ecdbbb47-8db2-47b3-ab81-edc4bfaa71bc",
    "mime_type": "text/plain",
    "name": "messages.log",
    "type": "file",
...},
{
    "id": "artifact--ecdbbb47-8db2-47b3-ab81-edc4bfaa71bc",
    "payload_bin": "binary_blob of last seen messages",
    "type": "artifact",
...}
```

Figure 26 STIX Format: Data and Attachments Example

This *indicator*, as exemplarily shown in Figure 27, contains concrete information on the intrusion, such as a textual description of the incident, the related types and incident classes and potential information on which pattern or rules exactly were violated.

```
{
    "confidence": 83,
    "created": "2020-01-26T17:55:10.442Z",
    "description": "For CAN ID 768 values can range from 0 to 255",
    "id": "indicator--9299f726-ce06-492e-8472-2b52ccb53191",
    "indicator_types": [
        "malicious-activity"
    ],
    "modified": "2020-01-26T17:55:10.442Z",
    "name": "Invalid CAN Payload",
    "pattern": "[x-can:id = 768 AND x-can:value >= 0 AND x-can:value <= 255]",
    "pattern_type": "stix",
    "pattern_version": "2.1",
    "spec_version": "2.1",
    "type": "indicator",
    "valid_from": "2021-02-25T10:31:20.588827Z"
}
```

Figure 27 STIX Format: Intrusion Indicator Example

In addition to that, it is possible to include complete log files or excerpts from the communication between components with the intrusion reporting message in the form of binary blobs with additional type specification or any other formatted text, such as CSV or JSON.

A.2.3 Examples of EARNEST Reports in STIX Data Format

ECUs that generate cross-partition frames are challenged by EARNEST and failed challenges are reported using as example the following STIX format.

```
{
    "id": "bundle--c6d5ea7d-2594-47cc-b3e9-8a0e1e8d8084",
    "objects": [
        {
            "_comment": "Meaning that the ECU has failed 100% of the challanges",
            "confidence": 100,
            "created": "2021-04-26T11:50:10.442Z",
            "_comment": "Meaning that the ECU has failed the challange when trying to send a frame with BO_ id 1170",
            "description": "Challange failed for BO_ id 1170",
            "id": "indicator--9299f726-ce06-492e-8472-2b52ccb53191",
            "indicator_types": [
                "malicious-activity"
            ],
            "modified": "2021-04-26T11:50:10.442Z",
            "name": "Invalid Challange",
            "spec_version": "2.1",
            "type": "indicator",
            "valid_from": "2021-02-25T10:31:20.588827Z"
        },
        {
            "created": "2021-04-26T11:50:10.442Z",
            "id": "identity--5206ba14-478f-4b0b-9a48-395f690c20a2",
            "identity_class": "system",
            "modified": "2021-04-26T11:50:10.442Z",
            "_comment": "A cross-partition frame is sent from untrusted zone to a trusted one of the in-vehicle network",
            "name": "CAN bus IPS for cross-partition frames",
            "roles": [
                "Cyber Security"
            ],
            "sectors": [
                "technology"
            ],
            "spec_version": "2.1",
            "type": "identity"
```


Figure 28: STIX object

Line 6 and 9 of Figure 28 represent a relevant part of the object in which the confidence of the malicious activity is reported. In particular, line 6 indicates the percentage of failed challenges reported by EARNEST, and line 9 shows the BO_ID, e.g., the CAN bus frame ID, used in the cross-partition frame.


```
        {
            "type": "observed-data",
            "spec_version": "2.1",
            "id": "observed-data--5046dbb6-0e6a-44bd-8b23-34ce41fae20a",
            "created_by_ref": "identity--f431f809-377b-45e0-aa1c-6a4751cae5ff",
            "created": "2021-04-26T11:50:10.442Z",
            "modified": "2021-04-26T11:50:10.442Z",
            "first_observed": "2021-04-26T11:50:04.432Z",
            "last_observed": "2021-04-26T11:50:05.442Z",
            "_comment": "Meaning that the ECU has failed 92 challanges",
            "number_observed": 92,
            "object_refs": [
                "artifact--ecdbbb47-8db2-47b3-ab81-edc4bfaa71bc"
            ]
        },
```


Figure 29: STIX object observed data

Always in the same object, in Figure 29 lines 59 and 60 indicate the observation time-window of the frame received by EARNEST. Then, in line 62, it is specified the number of challenges done that are consequence of the same number of cross-partition frames sent using that CAN id as shown in Figure 28, i.e., 1170.