IoT4CPS – Trustworthy IoT for CPS

FFG - ICT of the Future
Project No. 863129

Deliverable D5.4.1
Identity, Security and Safety in Product Lifecycle Data Management

The IoT4CPS Consortium:
AIT – Austrian Institute of Technology GmbH
AVL – AVL List GmbH
DUK – Donau-Universität Krems
IFAT – Infineon Technologies Austria AG
JKU – JK Universität Linz / Institute for Pervasive Computing
JR – Joanneum Research Forschungsgesellschaft mbH
NOKIA – Nokia Solutions and Networks Österreich GmbH
NXP – NXP Semiconductors Austria GmbH
SBA – SBA Research GmbH
SRFG – Salzburg Research Forschungsgesellschaft
SCCH – Software Competence Center Hagenberg GmbH
SAGÖ – Siemens AG Österreich
TTTech – TTTech Computertechnik AG
IAIK – TU Graz / Institute for Applied Information Processing and Communications
ITI – TU Graz / Institute for Technical Informatics
TUW – TU Wien / Institute of Computer Engineering
XNET – X-Net Services GmbH

For more information on this document or the IoT4CPS project, please contact:
Mario Drobics, AIT Austrian Institute of Technology, mario.drobics@ait.ac.at

Document Control

Title: Identity, Security and Safety in Product Lifecycle Data Management
Type: public
Editor(s): Violeta Damjanovic-Behrendt
E-mail: violeta.damjanovic@salzburgresearch.at
Author(s): Violeta Damjanovic-Behrendt (SRFG), Heribert Vallant (JR), Kai Nahrgang (JR)
Doc ID: D5.4.1

Amendment History

Version	Date	Author	Description/Comments
V0.1	01.06.2019	Violeta Damjanovic-Behrendt	Document organization
V0.2	15.07.2019	Violeta Damjanovic-Behrendt	The document version sent out to the partners: JR, TUG-ITI, TTTech, XNET
V0.3	14.08.2019	Heribert Vallant, Kai Nahrgang	Description of threats added
V0.4	25.08.2019	Violeta Damjanovic-Behrendt	Final draft of the report
V0.5	09.09.2019	Violeta Damjanovic-Behrendt	The report sent for the QA
V0.6	10.09.2019	Silvio Stern	Internal QA
V0.7	19.09.2019	Christoph Schmittner	Internal QA
V0.8	20.09.2019	Mario Drobics	Internal QA
V1.0	03.10.2019	Violeta Damjanovic-Behrendt	Final version of the report

The IoT4CPS project is partially funded by the "ICT of the Future" Program of the FFG and the BMVIT.

Federal Ministry
Republic of Austria
Transport, Innovation
and Technology

FFG
Forschung wirkt.

Content

Abbreviations

ABS	Anti-lock Braking System
ADAS	Advanced Driver Assistance System
ADS	Automated Driving Systems
CAD	Computer Aided Design
CAM	Computer Aided Manufacturing
CC	Cyclomatic Complexity
CIM	Computer Integrated Manufacturing
CITS	Cooperative Intelligent Transport System
CPS	Cyber Physical System
CSRF	Cross-Site Request Forgery
CTI	Cybersecurity Threat Intelligence
CySiVuS	Cybersecurity for Transport Infrastructure and Road Operators
DDoS	Distributed Denial of Service
DIF	Decentralized Identity Foundation
ECU	Engine Control Unit
GDPR	General Data Protection Regulation
HIDS	Host-based Intrusion Detection System
IDM	Identity Management
IDP	Identity Provider
IoT	Internet of Things
ITS	Intelligent Transport Systems
LOC	Lines of Code
ML	Machine Learning
NASAMDP	NASA Metric Data Program
NIST	National Institute for Standards and Technology
NREN	National Research and Education Networks
OBD	OnBoard Diagnostics
PAT	Pangea Arbitration Token
PII	Personally Identifiable Information
PLCDM	Product LifeCycle Data Management
SAML	Security Assertion Markup Language
SFC	Software Fault Prediction
SP	Service Provider
SPI	Sensitive Personal Information
SSO	Single Sign-On
V2I	Vehicle-to-Infrastructure
V2V	Vehicle-to-Vehicle
XML	eXtensible Markup Language
XSRF	Cross-Site Request Forgery

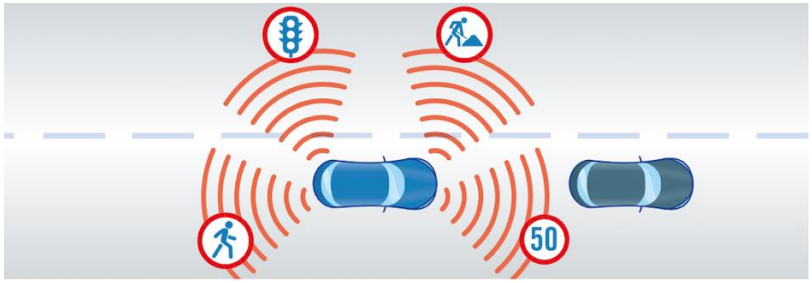
Executive Summary

This report captures identity, security and safety aspects of two automotive manufacturing and automotive driving scenarios, towards the definition of an extended data model covering the entire lifecycle of connected cars. In this report, we enhance the data model created in D5.2 “Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives” by integrating the security and safety features based on the use cases (from D5.2) and threats identified in WP4 “Security Verification and Analysis” of IoT4CPS. The resulting, extended data model ensures the inclusion of both multi-stakeholder and IoT-/ CPS-based assets (and their services) over lifecycle phases of the connected cars, and adds the third cybersecurity perspective to it. With such a model, our aim is to enable “digital twinning” of the real-world situations and processes related to lifecycle phases in the Automotive Driving and automotive Smart Manufacturing sectors, emphasizing the importance of a wide range of automotive safety and security indicators.

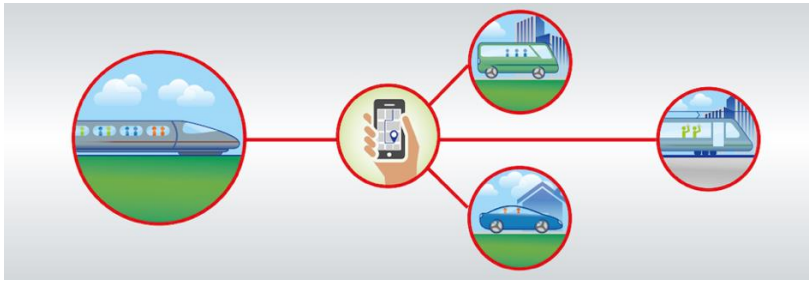
1. Introduction

The predecessor report D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives" captures multi-tenancy aspects related to smart and connected cars, actual legislations and emerging standards for data and information exchange in the Automotive Industry, along the entire product lifecycle. The resulting data modelling strategy, presented in D5.2, considers the roles of various stakeholders involved in PLCDM. The aim of this report D5.4 "Identity, Security and Safety in Product Lifecycle Data Management" is to further extend the data model from D5.2 by addressing identity, security and safety indicators, along the same product lifecycle. Here, we discuss the two use cases presented in D5.2 that combine the Device.CONNECT™ business case (defined in the IoT4CPS project by the partner AVL) and the following two use cases on Automotive Mobility (presented in the "Austrian Action Programme on Automated Mobility" (BMVIT, 2019)):

- "Safety+ through an all-round view": This use case is about driver assistance systems that use predictive sensors to intervene in traffic situations whenever danger is imminent. The information from other road users and from the infrastructure itself benefits to this use case, by enhancing road safety in the immediate environment of the vehicle.



- "New flexibility": This use case is about automated vehicles that offer new, on-demand services that can increase the flexibility of mobility users (e.g. route optimization, driving times tailored to personal preferences, secure and convenient connection mobility with intermodal transfer points, booking services, etc.) and ease the burden on the environment (e.g. by decreasing the environmental impact).



Practically, the above two use cases from (BMVIT, 2019) are extended to address the specific requirements of IoT4CPS and are defined in D5.2 as "Safety & Cybersecurity+ through the Lifecycle Stages" and "Assistive Intelligence+ through the Lifecycle Stages". Section 2 starts with the discussion on user identity and device identity in the cloud, and continues with a classification of major Identity Management (IDM) systems, including novel blockchain-based IDMs. Section 2 further discusses the importance of security and safety features in the smart car ecosystem. Section 3 creates the relations to other security tasks in the project and to the AVL's business case on security verifications along the product lifecycle. In Section 4, we provide the cybersecurity and safety analysis of two automated mobility use cases in the sense of their PLCDM. Such analysis includes identification of assets and stakeholders involved in the use cases, and for each of them, it

provides relevant identity, security, safety, and privacy risks. The analysis is a basis for the data acquisition (we looked at public open datasets and repositories) and further implementation of the Digital Twin prototype analytics in task T5.5. Section 5 lists relevant security and safety threats defined in WP4 of IoT4CPS. Section 6 lists relevant public security datasets, which are used in Section 7 to extend previously defined data model (see D5.2) in order to address cybersecurity and safety indicators in PLCDM. Section 8 concludes this report.

2. Cybersecurity Features in Product Lifecycle Data Management

2.1 User Identity and Device Identity

User identity and device identity management are both integral parts and enablers of IoT-based product lifecycle data management (PLCDM). The Automotive Driving applications and their smart assets (sensors, things, devices, powertrain controls, chassis controls, infotainment, communication devices, diagnostic and maintenance systems and tools, etc.) are based on modern Internet of Things (IoT), Cyber Physical System (CPS) and cloud technologies characterized by the increased connectivity and number of interfaces and thus, are exposed to a vast attack surface. On the one hand, smart applications and their assets are designed to improve information sharing amongst stakeholders, and on the other hand, they can enable implicit share of privacy data (e.g. personal data, personal information, Personally Identifiable Information (PII), or Sensitive Personal Information (SPI)) through massive surveillance, allowing malicious entities to modify and control data in a way that can threaten the privacy and safety of all stakeholders in the smart car ecosystem. The safety features of the car could be overridden by hackers exploiting vulnerable security flows, algorithms and relevant software technologies designed to improve the effectiveness of safety of Automated Driving Systems (ADSs) for e.g. collision prevention and mitigation (pedestrian detection, Anti-lock Braking System (ABS), obstacle detection, etc.), braking, hill assisting, terrain and wheel information sensors (temperature warning, tire pressure monitoring system, etc.), and more. Hence, the user identity and the device identity management, access controls and governance mechanisms need to be designed to jointly control information sharing and need to be considered early stage. For example, an identity lifecycle begins when the user activates a new device, creates an account and logs in for the first time. The user identity and privacy aspects continue to be exchanged through the entire product lifecycle, every time when the user information is exposed through diverse business and stakeholders' identity access systems.

User identity in product lifecycle data management controls the entire process of user's access to data, which needs to be regulated through designed login procedures and defined authorization levels for access controls. In complex cloud-based business systems with thousands of users, identity management requires consistent governance mechanisms and protocols for access management. For example, the user identities of terminated accounts related to end-of-life phase of the car, need to be removed from the car and from the cloud data centres, or anonymized and protected from possible manipulations in the future. Furthermore, the applications and services linked to the removed user identities need to be automatically reset to factory settings.

In addition to user identity, we also look at the device identity in product lifecycle data management. Adding new IoT products, services and devices onto networks is an easy step in comparison to their continuous functionality control and maintenance. In IoT4CPS, our approach is to take the advantage of networked devices, collect relevant data about the usage, maintenance

and functionality of these devices, create a Digital Twin-based prototype of these physical objects/devices that is fed by sensors and other data sources, and finally, perform related security, safety and privacy tests directly on a Digital Twin, based on data gathered from the system itself. Using a Digital Twin as a concept to design and implement our analytical prototype, we aim to support an effective notification mechanism that reacts to a variety of problems occurring on the system. We also aim to enable more advanced analytics to identify the cause of the problems and suggest how they can be efficiently fixed when the system is operating remotely and without a maintenance team accessing the system. Above all, our Digital Twin-based prototype is designed to enable identity, security and safety tasks to become a continuous process through constant monitoring of the system and its continual cybersecurity improvement.

Both the user identity and device identity management enable a source of trust for cloud-based services for all tasks related to authentication and authorization. They support a scheme of managed settings and authentication for both users and IoT devices. For example, the device identity manager uses specific authentication mechanism to identify the user linked to the device via the device manager. Usually, the user enters a password only during the initial setup phase and needs to set only personal preferences, e.g. notifications, working modes, etc. Sometimes, the device identity manager requires only user's email address to be specified, or could be set to automatically fill the URL field.

Finally, to discover already established networked system and identify devices and applications on the system, an open source tool called NMAP (for more details, see https://nmap.org/) can be used to assist in mapping out the network of devices and security auditing. Other useful network scanning tools are SuperScan from Foundstone, or NetScanTools Pro from Northwest Performance Software, both supporting automatic device discovery and identification.

2.1.1 Classification of IDentity Management (IDM) Systems

Recent research directions on IDM refer to the following four models that differ in scalability, privacy and user controls (Selvanathan et al., 2019) (Zwattendorfer et al., 2014):

- The isolated identity model is the simplest, traditional IDM that merges the Service Provider (SP) and the IDentity Provider (IDP), and allows for identification and authentication to be carried out at the SP. In this model, in order to access services of another SP, the user needs to register at the other SP's IDM. Managing of the diversity of credentials for accessing various service providers may become difficult for users (Jøsang and Pope, 2005).
- The central identity model allows the IDP to take over all identity-related services for the SP, e.g. identification and authentication, the management of the identity lifecycle, etc. (Bertino and Takahashi, 2011). In this model, the users' identity data are stored in a central repository at the IDP and the SPs do not need to maintain identity data in their own repositories (Cao and Yang, 2010). Some examples of the central IDM models are Kerberos (Neuman et al., 2005) and the Central Authentication Service (CAS).
- The user-centric identity model stores all identity data directly in the user's domain, i.e. on a secure token, a smart card, etc. In this model, the user's identity data can be transferred by an IDP to an SP only after the user gives the consent, which tremendously increases users' privacy (Dabrowski and Pacyna, 2008). Some examples of the user-centric model are the Windows CardSpace project and several national eID solutions such as the Austrian citizen card (Leitold et al., 2002), the German eID (Fromm and Hoepner, 2011), etc. One of the major concepts of the

Windows CardSpace was an InfoCard, a collection of claims about an identity that could be chosen during an authentication request or switched during communication with an SP.

- The federated identity model stores the identity data in a distributed manner, across different IDPs and/or SPs. In this model, there is no a single entity that is in full control of the identity information (Palfrey and Gasser, 2007). All IDPs and SPs form a federation and share a common trust relationship amongst each other, which is usually established on an organizational level; the enforcement is carried out through the platform, on a technical level. This model supports identification and authentication across different domains, enabling cross-domain Single Sign-On (SSO) (Cao and Yang, 2010). Popular examples of this approach are the Security Assertion Markup Language (SAML), Shibboleth, or WS-Federation (Kaler and McIntosh, 2009). SAML 2.0 is an OASIS standard that provides an XML-based framework for creating and exchanging security information between the users (for more details: http://saml.xml.org/). The Shibboleth project started as an Internet2 Middleware Initiative in 1999, and was focused on the development of interoperable identity and access management between web-based resources. Shibboleth is an implementation of the SAML protocol that shows excellent scaling, both in performance and manageability, and can be extended to support custom scenarios (for more details: https://www.shibboleth.net/products/identity-provider/). Some other examples of the federated identity models are the federations operated by various National Research and Education Networks (NREN), e.g. IDEM (www.idem.garr.it) by the Italian NREN - GARR, AAF (aaf.edu.au) by the Australian NREN - AARNET, and eduIDM (www.eduidm.ma) by the Moroccan NREN-MARWAN (Haddouti and Kettan, 2015).

Research on cloud-based IDM differentiates between the following identity models (Zwattendorfer et al., 2014):

- the identity IN the cloud model is similar to the isolated identity model in which users' identity data are stored in the domain of the cloud SP. The model minimizes the control of users over their data in the cloud (e.g. Google, Salesforce.com), but does not support a simplified login process (i.e. SSO);
- the identity TO the cloud model is similar to the central identity model. In this model, SP and its applications are cloud-based, whereas the IDP is not deployed in the cloud and data are not disclosed to a cloud SP. The IDP transfers identity and authentication data to the cloud SP through standardized interfaces that are based on SAML, OpenID, OAuth.
- the identity FROM the cloud model is also known as "Identity as a Service Model" (Ates et al., 2011). In this model, both the cloud application and the IDP are operated in the cloud by cloud SPs. Some examples of this model are Google Accounts Authentication and Authorization, or Facebook Login.
- the cloud identity broker model is an extension of "the identity FROM the cloud model", in which the IDP acts as a cloud identity broker or a hub between one or more SPs and one or more IDPs (Cloud Security Alliance, 2011) (Huang et al., 2010) (Zwattendorfer et al., 2013). By introducing the broker concept, this model hides the complexity of the individual IDPs from the SP. In addition, the SP needs to implement only one interface (to the identity broker). Here, there is only one strong trust relationship that is required between the SP and the identity broker. Some relevant examples of this model are McAfee Cloud SSO, the SkIDentity, the Cloud ID Broker, etc.

- the federated cloud identity broker model combines the traditional federated identity model with the cloud identity broker model (Zwattendorfer et al., 2013). In this model, the user and SPs can rely on the individual broker of their choice, which eliminates the drawback of being dependent on the same identity broker.
- the BlindIdM model is another extension of "the identity FROM the cloud model" (Nunez et al., 2012) (Nunez and Agudo, 2014). This model enables identity data storage and data processing to be performed by the semi-trusted IDPs in the cloud, or without knowing the contents of these data. Hence, the IDP provides these data in a blind manner (Nunez and Agudo, 2014), by using a proxy re-encryption scheme (Green and Ateniese, 2007) (Ateniese et al., 2006). This is an innovative contribution to the identity management solutions.
- the privacy preserving federated cloud identity broker model improves privacy preservation for users, by combining the advantages of the "federated cloud identity broker model" with the advantages of the BlindIdM model. This model can be used with semi-trusted cloud identity brokers (Zwattendorfer, 2014). It requires two re-encryption steps, since identity data flow through at least two cloud identity brokers. In addition, the user has to generate two re-encryption keys (one for the direction Identity Broker 1 → Identity Broker 2 and another one for the direction Identity Broker 2 → SP), which makes this model complex to implement. The implementation also requires an appropriate governance model to be put in place, to support the use of proxy re-encryption.

Storing digital identities on servers or in the cloud is still a considerable cybersecurity threat that will not retreat on its own. According to the 2019 MidYear QuickView Data Breach Report[1], only the first six months of 2019 have brought more than 3,800 publicly disclosed breaches that involved about 4.1 billion compromised records. The emergence of digital identity and IDM systems based on blockchain technology allows for the new model of identity, known as Self-Sovereign Identity. It enables everyone in the network (identity owners) to have the same source of truth about the validity of credentials (identity issuers) and identity verifiers, who attested the validity of the data inside the credentials and without revealing the actual data. It reduces, or even removes the need for inclusion of third parties, Blockchain based IDM systems provide security, transparency and ease-of-use when creating and managing digital identities that are stored on a decentralized system.

The recent public guidance on blockchain in identity management in the European Union is presented in (Lyons et al., 2018), (Lyons et al., 2019), both published by the European Union Blockchain Observatory & Forum. A set of standards already emerges to support blockchain-based IDM systems, including:

- Decentralized Identifiers and Verifiable Credentials, from the World Wide Web Consortia (W3C)
- Open Badges, from Mozilla and IMS Global, and
- Universal Resolver and Identity Hubs, from the Decentralized Identity Foundation (DIF).

NIST (National Institute of Standards and Technology) published a white paper that categorizes blockchain-based IDM systems into a taxonomy based on differences in 32 architecture,

[1] 2019 MidYear QuickView Data Breach Report. Online: https://pages.riskbasedsecurity.com/2019-midyear-data-breach-quickview-report

governance models, and other salient features (NIST Blockchain, 2019). Finally, there are many projects working towards blockchain adoption, at present (Kariuki, 2019):

- Aetna[2] for creating and verifying patient eligibility without a medical ID card;
- Authenteq[3] that uses AI-driven automatic ID verification;
- Bitnation[4] that features a blockchain-based form of ID and a public notary service;
- Blockverify[5] for monitoring product supply chains;
- BlockAuth[6] for identity verification and authentication;
- Blockstack[7] for user-controlled identity and peer-to-peer payments;
- Civic[8] for identity verification and authentication,
- Danube[9] supports customers to create, manage, use, destroy, secure and share their online identities on their own terms, without delegating the power to intermediaries to do this;
- DIF (Decentralized Identity Foundation)[10].

Other blockchain-based ID solutions are EverID, Evernym, Gov.UK Verify, Hyperledger, Idemia, IDKeep, IRMA, Key tokens, SelfKey, LifeID Foundations, Mooti, Netki, OneID, OpenID, PAT (Pangea Arbitration Token), Shocard, Thoreon, UniquID, uPort, Verses One, XID, and more (Kariuki, 2019). Another comparative study of IDM method using blockchain technology is provided in (Nabi, 2017). Apart products mentioned in (Kariuki, 2019), the authors in (Nabi, 2017) discuss also Cambridge Blockchain LLC, CredyCo, ExistanceID, Guardtimeas BLT, HYPR, OIXNet, etc.

Securing IoT and CPS using blockchain technology is based on public key cryptography for device IDM that substitutes default login credentials for devices. In addition, firmware installations on IoT devices will be possible only for the manufacturers signing the digital content using their private key stored on a blockchain, which will reduce security risks and facilitate user authentication and device verification.

2.2 Security and Safety

After identifying users (stakeholders) and IoT/CPS devices (and their services) in the system, the next step in our research is to perform relevant security and safety assessments. This phase starts with the prioritization of tests, based on potential risks related to each device/asset and criticality of their services in the system. For example, it makes sense to start with security and safety assessment of those assets with highest vulnerability (e.g. network exposure) or largest potential risk (e.g. drive control). The objective of such assessments is to examine all assets involved in corporate processes, gather detailed information about them and eventually, find associated

[2] https://www.aetna.com/
[3] https://authenteq.com/
[4] https://tse.bitnation.co/
[5] http://www.blockverify.io
[6] http://blockauth.org/
[7] https://blockstack.org/
[8] https://www.civic.com/
[9] https://danubetech.com/
[10] https://identity.foundation/

vulnerabilities. Section 5 of this report lists 8 security and safety threats based on the AVL's Device Connect Framework in IoT4CPS, and selected from the relevant report in WP4.

The identified vulnerabilities need to be mitigated in order to protect the system. This phase includes configuring and updating each asset to strengthen its security and comply with corporate governance models and the security standards. Once the proper security measures are established, the next phase of the security and safety lifecycle is to continue with relevant monitoring procedures, in order to ensure that desired security and safety posture of the system remains in place. Furthermore, the system needs to be monitored for intruders through an IDS system, or monitored for changes to identify any new vulnerabilities caused by newly installed applications or missing security patches, or monitored to ensure that the integrity of the system is maintained even when it is used by the authorized users.

Figure 1 illustrates major cybersecurity phases related to lifecycle data management, starting from traditional security methods for assessment and detection of vulnerabilities, through continuous security monitoring and advanced Cybersecurity Threat Intelligence (CTI) based on data analytics, visualization and dashboards. The purpose of the advanced security and safety analyses is also to support autonomous security responses and recovery, and automate security dissemination, e.g. through the use of the world-wide security repositories and platforms for sharing cybersecurity indicators (see https://www.misp-project.org/).

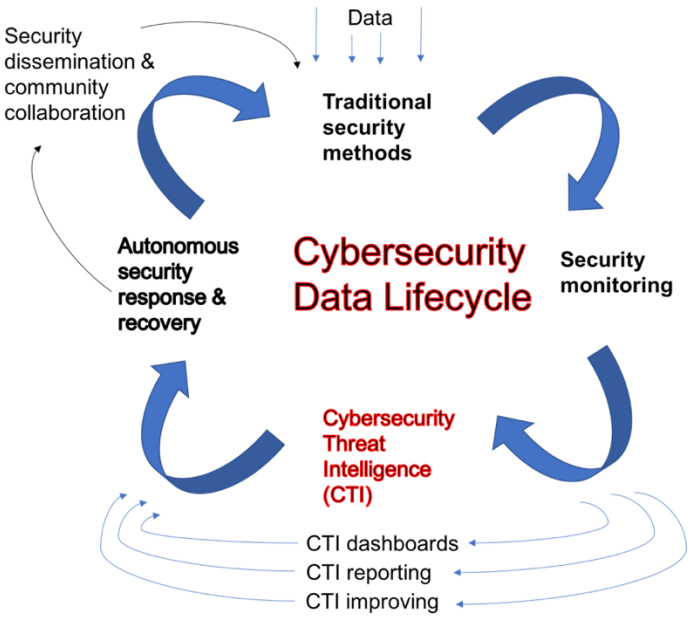


Figure 1 – Cybersecurity data lifecycle with advanced Cybersecurity Threat Intelligence and automated dissemination steps

3. Relation to Other Tasks in IoT4CPS

This report is closely linked to D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives" that discusses multi-tenancy aspects of the Automotive Driving applications along the entire product lifecycle. This report further extends the data model defined in D5.2, in order to include the desired identity, security and safety aspects of the Automotive Driving applications and services. Furthermore, this report looks at D2.2 and several tasks of WP3, WP4 and WP7.

3.1 Relation to IoT4CPS Business Case on Security Verification Along the Lifecycle (D2.2)

The IoT4CPS D2.2 "Business needs consolidation – competitive intelligence", Section 3.2 "AVL: Security Verification Along the Full Life Cycle of IoT-based Industrial Instrumentation Systems" reports on the role of the Device.CONNECT™ system, which enables communication links with the external systems through, e.g. smart/ predictive maintenance services in the cloud. It provides connectivity to a multitude of cloud-based commercial products, such as emission analysers, particle samplers, instrumentation systems, etc. At the same time, the cloud-based nature of the Device.CONNECT™ puts this device into a category of highly vulnerable and critical assets that need to be continuously monitored and checked against common threat intelligence indicators, regulatory compliance obligations and stakeholder's governance rules, in order to effectively responds to both cyber incidents and regulatory challenges.

In this report, we use the AVL's business case (the Device.CONNECT™ system) in the context of two selected use cases from (BMVIT, 2019). That way, we enhance the selected use cases to capture diversity of cloud-based identity, cybersecurity and safety issues and measures.

3.2 Relation to IoT4CPS work in WP3, WP4 and WP7

Figure 2 illustrates the major concepts in IoT4CPS, e.g. product lifecycle (PLCDM, in green), security aspects (in blue); trustworthy connectivity (in orange), and Digital Twin demonstrator (in pink). These concepts are related to other tasks and WPs in the following way:

1. Cybersecurity Lifecycle (joint work through WP3, WP4, WP5) that is based on data models created in tasks T5.2 and T5.4;
2. Digital Twin modelling (WP5) with the initial concepts and building blocks presented in D5.5.1 "Lifecycle Data Management Prototype I";
3. Trustworthy connectivity (WP7);
4. Traceability through lifecycle phases (WP7), related to task T7.2;
5. Security by isolation (WP7), related to task T7.3;
6. Smart production use case (Device.CONNECT™) (WP2, WP7) as described in section 3.1;
7. Autonomous vehicles (WP6), related to task T6.1 on secure and safe platform for Automated Driving applications.

Figure 2 also illustrates the Cybersecurity Data Lifecycle that adds identity, security and safety features to the main PLCDM observations, which are virtualized and implemented in the IoT4CPS Digital Twin prototype.

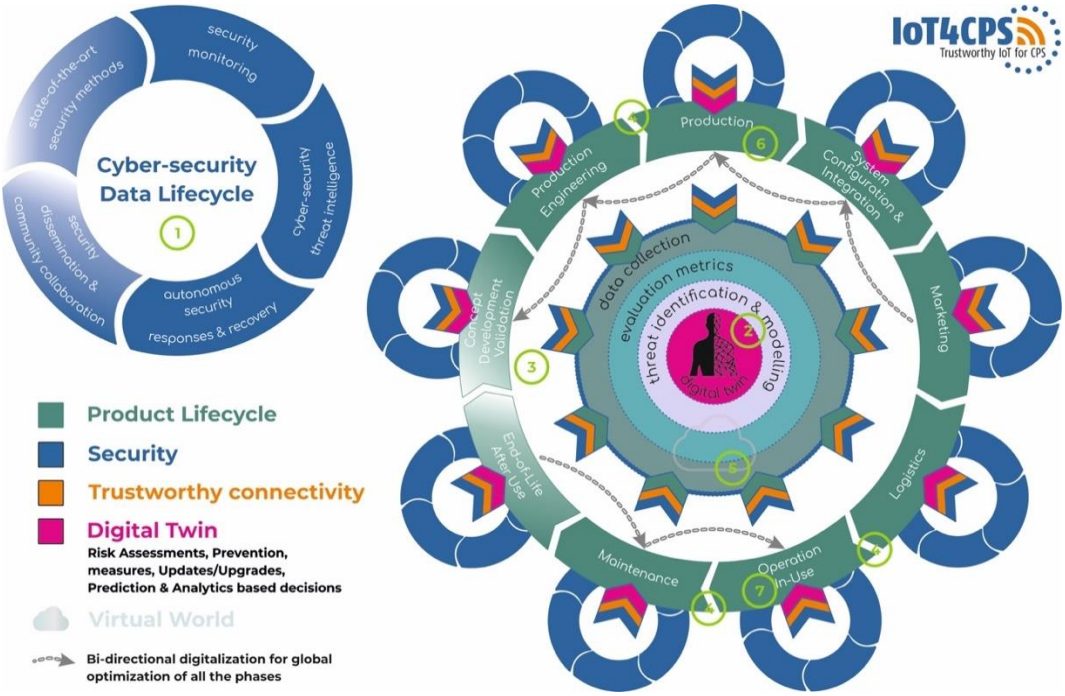


Figure 2 – The major concepts in IoT4CPS, in relation to the definition of the Digital Twin data models

4. Cybersecurity Analysis of Two Automated Mobility Use Cases and Their PLCDM

The Automotive Driving applications for connected cars are designed to assist the users in a variety of ways, from the enhancement of the driver's user experience to reducing distractions and improving the overall safety on the roads (ENISA, 2016). Such applications are beneficial to many stakeholders, e.g. smart car owners, drivers, passengers sharing the car, applications of smart cities and smart roads, insurance companies, environmental and climate change organisations, car sharing dealers, and many more. However, a cloud-based nature of such applications adds to their high exposure to a vast attack surface that can seriously compromise security, safety and privacy of the entire car ecosystem, e.g. from hijacking the steering wheel to changing registration numbers.

In 2015, security researchers for the first time demonstrated how easy is to hack a connected car by using simple techniques like password guessing (for more details, see the case of hijacking a Chrysler Jeep Cherokee during Black Hat 2015 security conference in Las Vegas (https://www.kaspersky.com/blog/blackhat-jeep-cherokee-hack-explained/9493/). In this specific case, researchers were able to completely control the music player, set the radio to whatever station they wanted and its volume to any level). In 2017, researchers from Trend Micro exploited Controller Area Network (CAN) protocols of connected vehicles using a physical hacking technique (https://blog.trendmicro.com/trendlabs-security-intelligence/connected-car-hack/). In 2018, researchers from the KU Leuven University in Belgium demonstrated how the key fobs signals can be used to open Tesla Model S vehicle's door "in a matter of second" (https://www.zdnet.com/article/how-to-steal-a-tesla-model-s-in-seconds/).

Figure 3 illustrates major modern car security risks related to Vehicle-to-Infrastructure (V2I), Vehicle-to-Vehicle (V2V), malware and spam, hacking of OnBoard Diagnostics (OBD) adapters (e.g. OBD-II) and/or car key fobs, third party apps, etc. Other relevant surveys on major obstacles to connected car uptake identify cybersecurity and privacy as the biggest concerns for the users (Levine, 2019) (Hitachi Systems Security, 2019). Similarly, Kaspersky team analysed connected car mobile apps for vulnerabilities and noted that none of the apps encrypt username and password credentials (see details here: https://securelist.com/mobile-apps-and-stealing-a-connected-car/77576/).

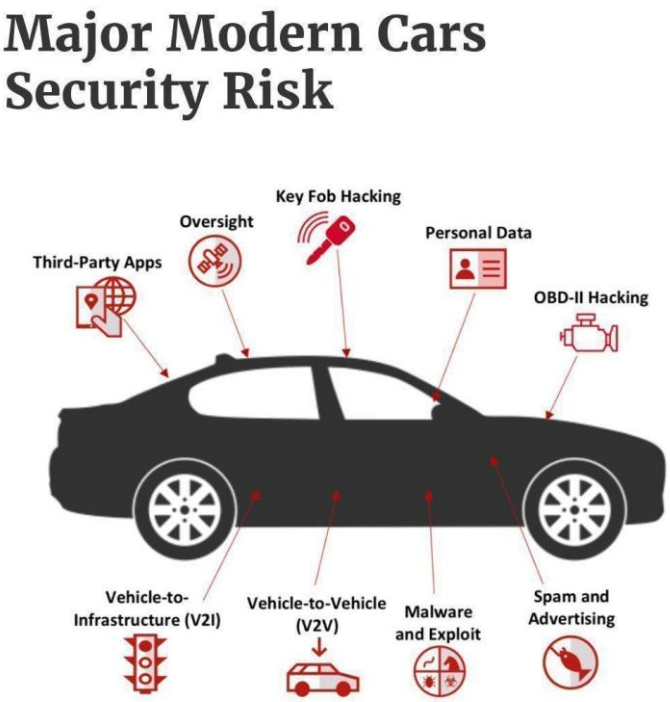


Photo source: https://www.itlab365.com/major-modern-cars-security-risk/

Figure 3 – The major modern cars security risks

The rest of this section provides cybersecurity analysis along the entire product lifecycle for the two selected use cases (for the detailed description of use cases, see D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives", Section 1.2):

- The use case "Safety & Cybersecurity+ through the Lifecycle Stages" includes (i) the Device.CONNECT™ for the data acquisition and management and (ii) its Digital Twin counterpart (demonstrator) that provides security and safety evaluations related to the Automotive Driving applications and verifications in IoT4CPS;
- The use case "Assistive Intelligence+ through the Lifecycle Stages" adds the Digital Twin demonstrator (as a counterpart of the Device.CONNECT™) and its analytics tools to enable additional assistive intelligence capabilities.

The following section analyses users (stakeholders) and devices (assets) for both above-mentioned use cases, and defines relevant identity, security, safety and privacy issues for each use case (see Table 1 – Table 4). The identification of both stakeholders and assets, and their assignment to relevant threat indicators through lifecycle phases is based on the analysis of two combined use cases and literature review on connected car security and safety features and recent incidents, e.g. (ENISA, 2016), (FPF, 2018) (Hitachi Systems Security, 2019) (Levine, 2019). In addition, the definition of relevant identity, security, safety and privacy issues for use cases is informed by the threat model that is defined in IoT4CPS D4.1 "Automotive Ethernet Protection Profile".

4.1 Analysis of the Use Case 1 "Safety & Cybersecurity+" to Capture Security & Safety Data Model Extensions related to PLCDM

Figure 4 illustrates the extension of the BMVIT's "Safety+ through an all-round view" by adding the AVL's Device.CONNECT™ system in IoT4CPS, in order to collect data related to the road and

environmental conditions, e.g. air pollution, temperature near the surface of the road, humidity. This data can be combined with the data from the car's powertrain and chasses controls. For example, the powertrain controls receive sensor information from electrical engines, transmission, wheels, etc. Chasses control receive sensor information related to both the car's frame and car's environment, including the steering and brakes, airbags, embedded cameras, real-view mirrors, windshield wipers (ENISA, 2016). The BMVIT's use case in Figure 4 is further extended to capture both cybersecurity and multi-stakeholder perspectives throughout the entire product lifecycle (including (i) initiation phase, (ii) operational phase, (iii) maintenance phase, and (iv) end-of-life phase). Note that sensor data collected through the use case assets (e.g. Device.CONNECT™, electrical engines, wheels or chasses controls, etc.) are all generated during the operational (driving) phase of PLCDM (see Figure 4). Our basic data strategy on how to capture missing datasets related to initiation phase (design, engineering, production), maintenance of the connected car and its end-of-life stage in IoT4CPS, is described in D5.2 "Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives".

Based on the cybersecurity analysis provided in this report, the same data model from D5.2 will be extended to additionally include specific cybersecurity datasets and threat indicators.

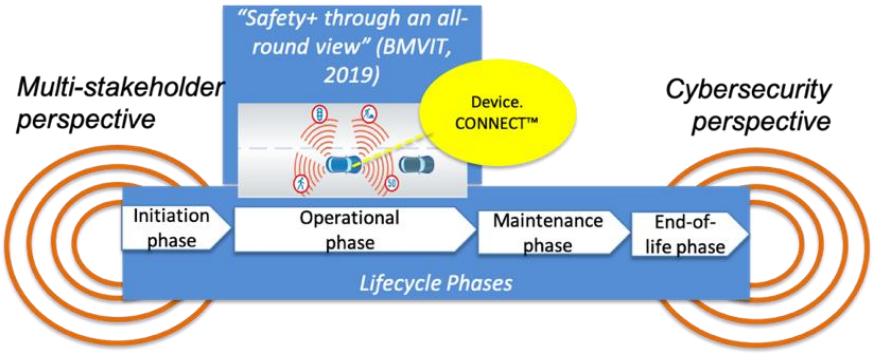


Figure 4 – Extension of the use case 1 to capture security & safety datasets and threat indicators related to various PLCDM phases

Identification of assets. Table 1 identifies various assets involved in the above illustrated use case. Furthermore, it assignes possible identity, security, privacy and safety issues related to each asset.

Table 1: Assets involved in the use case 1

Connected car's device/ sensor/ CPS	Type of sensor data	Relevant identity, security, privacy, safety issues
Initiation phase		
CAD uploader	Computer Aided Design (CAD). It assures that the design of the CPS-based product is analysed, optimized and sent for manufacturing.	Device identity theft; identity fraud. Physical harm of electrical or mechanical manufacturing processes caused by vulnerable behaviour of the system.
Collaborative analysis checker	It enables collaborative design and further improvements of CPS-based products to be manufactured.	Device identity theft; identity fraud. Provision of false design information leading to destruction of assets and safety issues.
CAM/CIM initiator	Computer Aided Manufacturing (CAM) / Computer Integrated Manufacturing (CIM). It enables the manufacturing flow from raw	Device identity theft; identity fraud. Provision of false information as a basis for CAM/CIM processes.

	materials to finished products, with quality assurance and automated assembly.	Safety issues related to automated assembly and quality assurance. Damaging effect on manufacturer reputation.
Robotic assembly checker	It checks the production of completed assemblies, part size, part defects (e.g. based on feeder jam data).	Device identity theft; identity fraud. Long term damage of manufacturing processes and assemblies.
Supply chain status control	It checks for the delivery terms in order to meet the demand.	Device identity theft; identity fraud. Provision of fake delivery details through malware injection, compromised digital signatures, etc. Access to sensitive corporate data and spying through backdoors installed on factory machines.
Operational phase		
Device.CONNECT™	Air pollution, temperature near the surface of the road, humidity data.	Device identity theft. Access to sensitive corporate data. Planting backdoors on corporate devices.
Powertrain control	Data from electrical engines, transmission data, wheels data.	Device identity theft. Remote control through hijacked sensors.
Chasses control	Data about the steering and brakes conditions, airbags, embedded cameras, real-view mirrors, windshield wipers.	Retrieving information about the vehicle, such as vehicle ID number, make, model, IP address, GPS coordinates. Scanning multiple mobile apps and connected devices to find out the owner of the vehicle in order to track a person. Wirelessly controlled radio stations, windshield wipers, air conditioning system, vehicle steering, etc. Compromised brakes, speed and gear controls.
Maintenance phase		
CIM Remote Monitoring Service	It monitors for unauthorized access and changes to the files and products (connected cars).	Device identity theft. Provision of false information.
Integrity Monitoring Service	It detects and reports changes made in files or detects manipulations.	Device identity theft. Provision of false information.
End-of-life phase		
Privacy Data Monitoring Service	It ensures that retained privacy data are removed from the connected cars (FPF, 2018), e.g. mobile apps that are used, mobile apps log-in data, location, the driver's daily route, phone contacts and address books, garage door codes, various digital content, subscription services, wi-fi hotspots, data services, etc.	Exploring privacy data stored in the connected car or in the cloud databases. Personal data sold or leaked to the public. Unauthorized access to privacy information.

Other Data Monitoring Services	It enables other data to be removed from the connected cars, e.g. OBD information.	Exploring other data in the connected car or in the cloud databases, e.g. electronic data recorders (black boxes), data collected from the vehicle monitoring devices, etc.

Identification of stakeholders. Sharing data amongst various stakeholders can open numerous privacy issues leading to reputational damage for the users, car manufacturers, suppliers, garages, network service providers, software and application providers, etc. Table 2 identifies the potential stakeholders that are (directly or indirectly) involved in the use case 1.

Table 2: Stakeholders involved in the use case 1 (partly based on (ENISA, 2016))

Stakeholder	Description of the stakeholder's role in the use case	Relevant identity, security, privacy, safety issues
Initiation phase:	Manufacturers & Suppliers	
Manufacturer	Provides the production and assembly of the car components.	Threatening safety and privacy of manufacturers. Reputational damage. Intellectual property theft.
Aftermarket Supplier	Provides components with additional features, e.g. media player.	Spying on corporate secrets. Reputational damage. Conflicting security and safety features.
Operational phase:	Car Users & Internal Services	
Driver	Drives and uses the connected car's gadgets and apps/services. Connects via smartphone. Uses external cloud applications.	Privacy and safety risks to the drivers. Personal data sold or leaked to the public.
Passengers	Use gadgets and apps or are exposed to apps and services running on other user's devices.	Privacy and safety risks to the passengers. Personal data sold or leaked to the public.
Powertrain control services	Transmission controls; wheels controls; services for monitoring of the engine features, etc.	Security and safety risks, e.g. a physical hacking technique to exploit the CAN protocol of a vehicle. Compromised and unexpected behaviour of cars, e.g. heating seats.
Operational phase:	External Services	
Road services	Monitoring road and traffic conditions; Safety recommendations and contextual insights, e.g. speed limit changes, roadway conditions. eCall services.	Disturbance of surrounding vehicles and road services. Safety issues through incorrect signalisation data or incorrect navigation data.
Testing and certification services	Monitoring driving habits; Contextual insights.	Privacy and safety risks to the drivers. Driver's disruption.
Insurance	Pay-How-You-Drive insurance plan.	Privacy risks and secrets.

services		Fraud situations. Unauthorized copies.
Network connectivity providers & services	Network access and services. Remote transmission of vehicle data. Remote engine start. Geo-fencing. Crash reporting and emergency warning (eCall), etc. Remote diagnostics and fleet management.	Integrity breach and disruptions. A loss of control of a car. Spoofed communication causing accidents.
Smart cities & services	Economical use of the road infrastructure. Smart city weather station and road speed controls. Environmental impact evaluation.	Trade secrets. Data confidentiality and privacy of citizens, drivers and passengers. Safety related vulnerabilities.
Maintenance :	External Services	
Road services	Monitoring traffic conditions; Safety recommendations.	Trade secrets. Safety risks.
Manufacturer	Evaluation of part's functionality and safety	Integrity breach and disruptions. Safety risks. Trade secrets.
End-of-life phase:	External Services	
Smart city services	Economical use of the city infrastructure. Environmental impact evaluation.	Security and safety vulnerabilities.

4.2 Analysis of the Use Case 2 "Assistive Intelligence+" for Security & Safety Data Model Extensions related to PLCDM

To support the assistive intelligence capabilities relevant to security and safety features, we "redefine" the use case on "New flexibility" (BMVIT, 2019) by adding the Device.CONNECT™ system and the Digital Twin demonstrator to automatically process data in a way that supports stakeholders along the entire lifecycle and verifies the system's safety and cybersecurity conditions (see Figure 5). The connected car collects data such as air pollution, temperature near the surface of the road, humidity, telematics data about braking, engine performance, collision detection and emergency calling, vehicle diagnostics, vehicle speed, GPS data. The power of data lies in its combination. For example, the connected car can recognize the intention of another car to change lanes, based on data related to the car's speed and position adjustment. Based on light signals, it can know which connected car will turn and which will continue straight. This type of scenarios shows a potential to eliminate traffic fatalities in the future.

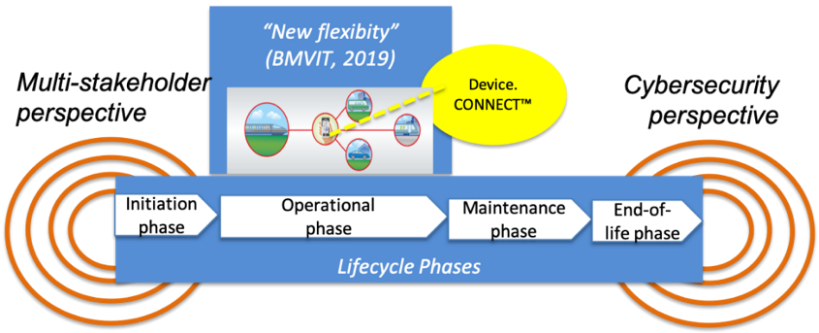


Figure 5 – Extension of the use case 2 to capture security & safety datasets and threat indicators related to various PLCDM phases

Identification of assets. Table 3 lists various assets involved in the use case 2. It assigns possible identity, security, privacy and safety issues related to each asset.

Table 3: Assets involved in the use case 2

Connected car's device/sensor/ CPS	Type of sensor data	Relevant identity, security, privacy, safety issues
Initiation phase		
Robotic assembly checker	It checks the production of completed assemblies, part size, part defects (e.g. based on feeder jam data)	Device identity theft. Access to sensitive manufacturing data. Provision of false assembly information that lead to safety issues.
Supply chain status control	It checks for the delivery terms in order to meet the demand	Device identity theft; identity fraud. Provision of fake delivery details through malware injection, compromised digital signatures, etc. Access to sensitive corporate data and spying through backdoors installed on factory machines.
Operational phase		
Device.CONNECT™	Air pollution, temperature near the surface of the road, humidity	Device identity theft. Access to sensitive corporate data. Planting backdoors on corporate devices.
Powertrain control	Data from electrical engines, transmission data, wheels data	Device identity theft. Incorrect data that lead the car to unsafe situations. Remote control through hijacked sensors.
Chasses control	Data about the steering and brakes conditions, airbags, embedded cameras, real-view mirrors, windshield wipers, Advanced Driver Assistance System (ADAS)	Device identity theft. Incorrect navigation and assistance data that lead the car to unsafe locations and situations. Retrieving information about the vehicle, such as vehicle ID number, make, model, IP address, GPS coordinates. Wirelessly controlled radio stations, windshield wipers, air conditioning system, vehicle steering, etc. Compromised brakes, speed and gear controls.
Infotainment control	Music and video streaming, Bluetooth connectivity, wi-fi connectivity and wi-fi hotspots, SMS texting …	Device identity theft. Scanning multiple mobile apps and connected devices to find out the owner of the vehicle in order to track a person.
External media	Mobile phones, Bluetooth speakers for cars, etc.	Device identity theft. Scanning multiple mobile apps and connected devices to find out the owner of the vehicle in order to track a person.

Maintenance phase		
Integrity Monitoring Service	It detects and reports changes made in files or detects manipulations	Device identity theft. Provision of false information.
End-of-life phase		
Privacy Data Monitoring Service	It ensures that retained privacy data are removed from the connected cars (FPF, 2018).	Exploring privacy data stored in the connected car or in the cloud databases. Personal data sold or leaked to the public.
Non-Privacy Data Monitoring Services	It enables other data to be removed from the connected cars, e.g. on-board diagnostic information.	Exploring other data in the connected car or in the cloud databases, e.g. electronic data recorders (black boxes), data collected from the vehicle monitoring devices, etc. Unauthorized access to information.

Identification of stakeholders. Table 4 identifies the stakeholders involved in the use case 2.

Table 4: Stakeholders involved in the use case 2 (partly based on (ENISA, 2016))

Stakeholder	Description of the stakeholder's role in the use case	Relevant identity, security, privacy, safety issues
Initiation phase:	Manufacturers & Suppliers	
Supplier	Provides car components and /or operating system for connecting car components.	Threatening safety and privacy of supplier. Reputational damage. Intellectual property theft.
Aftermarket Supplier	Provides components with additional features, e.g. media player.	Spying on corporate secrets. Reputational damage. Conflicting security and safety features.
Operational phase:	Car Users & Internal Services	
Driver	Drives and uses the connected car's gadgets and apps/services. Connects via smartphone. Uses external cloud applications.	Privacy and safety risks to the drivers. Personal data sold or leaked to the public.
Passengers	Use gadgets and apps or are exposed to apps and services running on other user's devices.	Privacy and safety risks to the passengers. Personal data sold or leaked to the public.
Cross-collaborative services and data exchanged among connected cars	Data received from another connected cars, e.g. the location of a car accident that another connected car spotted on the road, or received accident information from other cars, or smart city info services.	Compromised and unexpected behaviour of cars, e.g. heating seats. Safety related vulnerabilities, e.g. based on a disturbance of warning/direction lights.
Operational phase:	External Services	
Smart cities & services	Economical use of road infrastructure.	Trade secrets. Data confidentiality and privacy of citizens, drivers and passengers. Safety related vulnerabilities.
Road services	Monitoring road and traffic conditions; Safety recommendations and contextual	Fraud situations. Unauthorized copies.

	insights, e.g. speed limit changes, roadway conditions.	Unauthorized access to information.
Insurance services	Pay-How-You-Drive insurance plan.	Privacy risks and secrets. Unauthorized copies.
Energy/fuel services	Energy/fuel supply.	Trade secrets. Safety related vulnerabilities.
Marketing services	Monitoring driving habits and user's preferences to create personalized offers.	Trade secrets. Data confidentiality and privacy of citizens, drivers and passengers.
Maintenance:	External Services	
Insurance services	Pay-How-You-Drive insurance plan.	Privacy risks and secrets. Unauthorized copies.
Road services	Monitoring traffic conditions; Safety recommendations.	Trade secrets. Safety risks.
End-of-life phase:	External Services	
Smart city services	Environmental impact evaluation. Weather data.	Trade secrets. Data confidentiality and privacy of citizens, drivers and passengers. Safety related vulnerabilities.

5. Relevant Security and Safety Threats

This section presents eight security and safety threats out of 328 threats defined in the IoT4CPS project (see D4.1 “Automotive Ethernet Protection Profile” for more details). The selected threats presented in this section belong to different threat categories, e.g. Threat 1 refers to Cross-Site Request Forgery; Threats 108-112 are all about security manipulations at the level of vehicle data; Threat 113 describes a case of an attack at the level of network; Threat 135 covers an unauthorized access to privacy data stored in the connected car’s infotainment system.

5.1 Cross Site Request Forgery (Threat 1)

Cross-Site Request Forgery (CSRF or XSRF) is a type of attack in which an attacker forces a user's browser to make a forged request to a vulnerable site by exploiting an existing trust relationship between the browser and the vulnerable web site. In a simple scenario, a user is logged in to web site A using a cookie as a credential. The other browses to web site B. Web site B returns a page with a hidden form that posts to web site A. Since the browser will carry the user's cookie to web site A, web site B now can take any action on web site A, e.g. adding an admin to an account. The attack can be used to exploit any requests that the browser automatically authenticates, e.g. by session cookie, integrated authentication, IP whitelisting… The attack can be carried out in many ways such as by luring the victim to a site under control of the attacker, getting the user to click a link in a phishing email, or hacking a reputable web site that the victim will visit. The issue can only be resolved on the server side by requiring that all authenticated state-changing requests include an additional piece of secret payload (canary or CSRF token) which is known only to the legitimate web site and the browser and which is protected in transit through SSL/TLS.

5.2 Manipulate Vehicle Data - Illegal/Unauthorised Changes to Vehicle's Electronic ID (Threat 108)

Category: [IoT4CPS] Target of an attack on a vehicle

The vehicle identification number is the identifying code for an automobile and serves as the car's fingerprint. A change of this ID could have far reaching implications. On the one hand, wrong software updates could harm the whole system and could introduce several safety related issues. On the other hand, wrong identifier would disguise the real identity of the vehicle in case of car theft. Also, the creation of spare keys to gain physically access to the car after sniffing the cars identity number is aligned to this threat.

5.3 Manipulate Vehicle Data - Identity Fraud (Threat 109)

Category: [IoT4CPS] Target of an attack on a vehicle

This attack is performed by using the identity of e.g. the automotive service station ID without authorization to manipulate the setup of the vehicles Engine Control Unit (ECU).

5.4 Manipulate Vehicle Data - Circumvent Monitoring Systems (Threat 110)

Category: [IoT4CPS] Target of an attack on a vehicle

The “In Vehicle Monitoring System” enables the owner of the car or a third party to track the vehicle's location by collecting time-spatial data. This feature normally can be divided into active and semi-passive tracking. When a cellular network is available the tracking device will connect and

transmits data to a server. Otherwise the data will be stored internally and will be transmitted to the server later when the network becomes available again. In case of a stolen car the attacker may manipulate this data to hide the exact location of the car. Also, other attacks denying the presence of the car and the driver at a certain time and place could be the aim of such a manipulation.

In addition, the monitoring data reflects the driver's behaviour and can record sudden braking or harsh acceleration and speeding which might influence the insurance premiums. Besides that, the reduction of incidents on the road by controlling speed limits would also be influenced by such an attack.

5.5 Manipulate Vehicle Data - Manipulation of Driving Data (Threat 111)

Category: [IoT4CPS] Target of an attack on a vehicle

Driving data is generated based on operations performed by the driver of the vehicle. An attacker might change this data to get better insurance premiums, e.g. Pay-How-You-Drive. Since this data consist of geographic information, user behaviour and technical information about the car, an attacker who tries to manipulate the monitoring systems or the diagnostic data have to manipulate this dataset as well to blur their attack.

5.6 Manipulate Vehicle Data - Diagnostic Data (Threat 112)

Category: [IoT4CPS] Target of an attack on a vehicle

Valid diagnostic data is a crucial point to be able to track problems of the specific car as early as possible. It could also affect the development process if serious faults are detected which have to be eliminated during the production. The aim of an attack could be to hurt a specific person by not reporting correct diagnostic values and causing an accident, or to harm the car manufacturer.

5.7 Attack on Network - Vehicle Acting as a Botnet (Threat 113)

Category: [IoT4CPS] SmartHub used as a means to propagate an attack

A botnet is a collection of internet-connected devices. Each of these devices is running malicious software which can be triggered to run a collaborative attack (e.g. Distributed Denial of Service (DDoS)) against another internet devise.

5.8 Extract Data/Code - Unauthorized Access to Privacy Information (Threat 135)

Target of an attack on a vehicle

Based on the data collected a detailed driver profiling might be possible. Depending on the information collected by vehicle especially a combination of time, location and the direction of movement could comprise information about friends, co-workers and relatives. Information collected by the entertainment system and the hands-free car kit could reflect the stress level or the physical condition of the driver.

6. Relevant Public Security Datasets

Apart a variety of connected cars', smart roads and smart cities' conditions, the Digital Twin demonstrator in IoT4CPS requires high-quality cybersecurity training datasets to create meaningful Machine Learning (ML) models. In the following, we analyse several prominent quality cybersecurity data repositories and datasets that are publicly available for the research experimentations.

6.1.1 Security Data Repositories

Available from: http://www.secrepo.com/

SecRepo is a repository of samples of security related data: network datasets (network scanning, traffic, c99 shell traffic, logs, and more), malware datasets (ZeuS botnet binaries, VirusShare.com repository of malware samples, OP Cleaver binaries, etc.), system logs, failed SSH attempts, data from various honeypots (e.g. Amun http://amunhoney.sourceforge.net/ and Glastopf https://securityonline.info/glastopf-web-application-honeypot/), and many more.

Another source of useful ML material and datasets for cybersecurity is available here: https://github.com/jivoi/awesome-ml-for-cybersecurity#-datasets Some of the datasets listed here are:

- DARPA Intrusion Detection Data Sets: https://www.ll.mit.edu/ideval/data/
- Stratosphere IPS Data Sets: https://stratosphereips.org/category/dataset.html
- Open Data Sets: http://csr.lanl.gov/data/
- Data Capture from National Security Agency: http://www.westpoint.edu/crc/SitePages/DataSets.aspx
- Malicious URLs Data Sets: http://sysnet.ucsd.edu/projects/url/
- Multi-Source Cyber-Security Events: http://csr.lanl.gov/data/cyber1/
- KDD Cup 1999 Data: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html
- Web Attack Payloads: https://github.com/foospidy/payloads
- WAF Malicious Queries Data Sets: https://github.com/faizann24/Fwaf-Machine-Learning-driven-Web-Application-Firewall
- Malware Training Data Sets: https://github.com/marcoramilli/MalwareTrainingSets
- Aktaion Data Sets: https://github.com/jzadeh/Aktaion/tree/master/data
- CRIME Database from DeepEnd Research: https://www.dropbox.com/sh/7fo4efxhpenexqp/AADHnRKtL6qdzCdRlPmJpS8Aa/CRIME?dl=0
- Publicly available PCAP (packet capture) files http://www.netresec.com/?page=PcapFiles
- 2007 TREC Public Spam Corpus: https://plg.uwaterloo.ca/~gvcormac/treccorpus07/
- Drebin Android Malware Dataset https://www.sec.cs.tu-bs.de/~danarp/drebin/
- PhishingCorpus Dataset: https://monkey.org/~jose/phishing/
- EMBER: https://github.com/endgameinc/ember
- Vizsec Research: https://vizsec.org/data/
- SHERLOCK: http://bigdata.ise.bgu.ac.il/sherlock/index.html#/

6.1.2 The ADFA Intrusion Detection Datasets

The datasets are available from: https://www.unsw.adfa.edu.au/unsw-canberra-cyber/cybersecurity/ADFA-IDS-Datasets/

The ADFA Intrusion Detection Datasets (2013) are created for Host-based Intrusion Detection System (HIDS) evaluation (Creech, 2014) (Creech and Hu, 2013).

6.1.3 The Cyber Research Center Datasets - ITOC CDX (2009)

Available from: https://westpoint.edu/centers-and-research/cyber-research-center/data-sets
The datasets are available from:
https://drive.google.com/uc?id=0B0u9Tg7udaAXaUFHRFpQWjR0dW8&export=download

ITOC CDX (2009) dataset provides a comprehensive set of log data under ongoing "sophisticated" attacks.

6.1.4 The NSL-KDD Dataset

The datasets are available from: https://www.unb.ca/cic/datasets/nsl.html

A benchmark data set that help researchers compare different intrusion detection methods (Tavallaee et al., 2009).

6.1.5 DARPA Intrusion Detection Datasets

The datasets are available from: https://archive.ll.mit.edu/ideval/data/1998data.html
Intrusion Detection Attacks database: https://archive.ll.mit.edu/ideval/docs/attackDB.html

1998 DARPA Intrusion Detection Evaluation included an off-line evaluation and a real-time evaluation. The evaluations were based on a sample of network traffic and audit logs, and a larger sample of training data on network-based attacks.

6.1.6 Public Datasets on Software Metrics

Public datasets on software metrics are mainly created from open source projects and often provide static code metrics (Lines of Code (LOC), Cyclomatic Complexity by McCabe (CC), etc.) and bug information (Altinger, 2016). Most of the bug commit information has been extracted using the SZZ algorithm, which is introduced by Śliwerski et al. (2015) as an approach to identify bugs in a software repository (including GitHub repositories). The name of the SZZ algorithm is given after the initials of the three authors.

One of the first public available datasets has been released by the NASA metric data program (NASAMDP) (NASAMDP, 2004) (online available from: http://mpd.ivv.nasa.gov). The NASAMDP datasets contain software metrics collected at ten different projects within NASA flight software. Another dataset containing software engineering data is called PROMISE (online available from: http://promise.site.uottawa.ca/SERepository/). PROMISE is founded and administrated by the authors of Sayyad et al., (2005) and Menzies et al., (2005). It includes 60 projects usable for Software Fault Prediction (SFP). Similarly, Software-artefact Infrastructure Repository (SIR) published by Do et al., (2005) can be considered to be the first database on software bugs, containing 81 projects with a rather small code size ranging from 24 LOC to 8.570 LOC.

Practically, the only industrial available dataset for SFP has been released by NASAMDP and the PROMISE repository (Altinger, 2016). Menzies et al., (2005) shows how ML approaches can be used to build up defect prediction models. For example, the following ML algorithms: OneR, J48, and NB,

can be used to predict error prone software modules. The NASAMDP and the PROMISE repository of software engineering data can be used for the evaluation of the predictions.

Finally, Altinger (2015) contains datasets on automotive software repository that is publicly available from: http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/MSR_2015_dataset_automotive.zip

6.1.7 Public Datasets on Software Bugs and Defects

A collection of bug datasets is given in Table 7.

Table 5: Public datasets on software bugs and defects (based on (Altinger, 2016))

Reference to dataset	Created for	Hosting
Zimmerman et al., 2007	Eclipse 2.0, 2.1 and 3.0.	25.210 files with 25.585 defects
Kamei et al., 2008	Eclipse 3.0 and 3.1	9.726 Java files of whom 16,98% are marked as faulty
Herraiz et al., 2009	5000 open source projects	N/A
Mockus et al., 2009	GoogleCode and SourceForge	1398 projects with 207.904.557 files in total
D'Ambros et al., 2010	Eclipse JDT Core, Eclipse PDE UI, Equinox framework, Mylyn and Apache Lucene projects	It contains software consisting of 2.131 classes and containing 1.923 bug commits.
Jus et al., 2014	Software testing research	The initial commit contains 357 bugs on five open source Java projects ranging between 22.000 and 96.000 LOC.

7. Extension of Multi Stakeholder-Centered Data Model in IoT4CPS

The D5.2 “Product Lifecycle Data Management (PLCDM) Stakeholder Perspectives” report (see Section 6) presents the design method for multi-stakeholder-centered data model of connected cars. The presented model addresses actual legislations and emerging standards for data and information exchange in the Automotive Industry, and in addition, identifies major assets involved over the entire product lifecycle. In this report, we further extend the data model from D5.2 in order to associate identity, security and safety aspects to both multi-stakeholders and assets, over the same lifecycle phases as in D5.2. The resulting data model assimilates diverse datasets and is used as a basis for Digital Twin prototype for cybersecurity decision-making support. Note that the authors in (Schmittner et al., 2019) present an approach that emphasizes the dependency triple between *Stakeholders – Cooperative Intelligent Transport Services (C-ITS) - Risks* towards the definition of a comprehensive automotive cybersecurity reference architecture. The presented approach is the result of the ongoing Austrian research project CySiVuS (Cybersecurity for Transport Infrastructure and Road Operators). Similarly, the IoT4CPS model assimilates the following three perspectives: *Stakeholders – Assets (and Services) over Lifecycle of Connected Cars – Risks*. The assets of the connected cars that are of interest to our data modelling are IoT/ CPS supported devices and their services. With such a model, our aim is to extend the observation of the connected car to include not only its operational phase (e.g. as shown in (Schmittner et al., 2019)), but the engineering of the car, its maintenance and end-of-life management phases (see Figure 6). Figure 6 illustrates the extension of the data model from D5.2 to capture cybersecurity and safety features along lifecycle phases in the Automotive Industry (for both Automotive Driving and automotive Smart Manufacturing domains, and including assets relevant to lifecycle).

The current data model needs to be normalized and fused in order to identify relationships, trends and anomalies related to security and safety vulnerabilities in the car ecosystem. The data models of the next Digital Twin prototype releases will be presented in upcoming reports D5.5.2 “Lifecycle Data Management Prototype II” and D5.5.3 “Lifecycle Data Management Prototype III”.

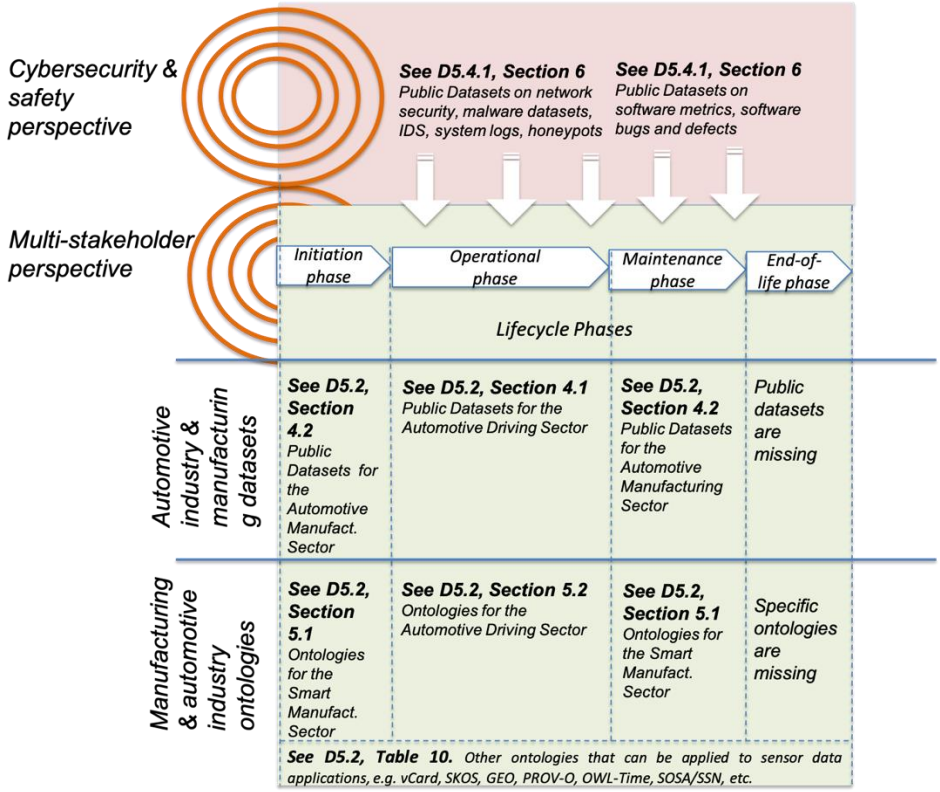


Figure 6 – Extension of the use case 2 to capture security & safety datasets and threat indicators related to various PLCDM phases

8. Conclusion

Identity, security and safety in the Automotive Industry have always been a concern for authorities, governance bodies, manufacturers and the public alike. However, a common standard allowing a complete integration of safety and security measures in the car PLCDM is still missing (ENISA, 2016). With the objective of making smart and connected cars more trustworthy, secure and safe, the IoT4CPS project addresses relevant cybersecurity threats and challenges, and experiments with promising technology solutions designed to offer preventive security measures and reduce safety accidents. The aim of this report is to provide a basis for quality data-centric Digital Twin prototype that can assimilate diverse datasets in order to support automated cybersecurity and privacy decision-making in the smart car ecosystem, along its entire product lifecycle. The design of such a prototype requires an effective data strategy to be put in place, and in parallel, it requires knowledge and understanding of policy and regulatory issues at national and international levels regarding smart car, cybersecurity and safety. For example, some General Data Protection Regulation (GDPR) data privacy fundamentals still need to be incorporated into smart car design, e.g. user's consent to share data.

To this end, several frameworks and best practices in the Automotive Industry are being developed to ensure smart car security-by-design. For example, ENISA (2016) categorizes the good practices as:

- Policy and standards, i.e. adherence to regulations, liability, traceability;
- Organizational measures, i.e. general measures, secure development and security until the end-of-life, and

- Technical good practices, i.e. communication protection; identification, authentication, authorization; security audit; self-protection; cryptography; user data protection.

In addition, the "Rolling Plan for ICT Standardisation 2019" that bridges between EU policies and standardisation activities in the field of ICT, points out at many requested actions related to automated driving, new reference data model for mobility services, security in the context of Intelligent Transportation Systems, advanced manufacturing, and more. A lot of thinking is still required about data management for rental cars and car manufacturers, e.g. about how the data uploaded from phones to infotainment systems of rental cars can be wiped off after the rental is over. Such data can contain information about driver's home address, or contents of the drivers and other passengers' smart phones, locations, daily route, credit cards, garage door codes, etc. Similarly, the procedures to erase the data upon car's discontinuing use, loss or sale of a car need to be standardized and transparent to users, e.g. by offering clear rules to erase the data and disconnect the car from other personal gadgets, or "factory reset" options for cleaning the data from the system.

9. References

(Altinger, 2015) Altinger, H., 2015. Dataset on automotive software repository, Feb. 26, 2015. Online available from: http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/MSR_2015_dataset_automotive.zip

(Altinger, 2016) Altinger, H., 2016. State of the Art Software Development in the Automotive Industry and Analysis upon Applicability of Software Fault Prediction. Doctoral Thesis. Graz University of Technology. 2016. http://www.ist.tugraz.at/_attach/Publish/AltingerHarald/PHD_Altinger_automotive_SW_analysis.pdf

(Ateniese et al., 2006) G. Ateniese, K. Fu, M. Green, and S. Hohenberger, 2006. Improved proxy reencryption schemes with applications to secure distributed storage. ACM Trans. Inf. Syst. Security, 9(1):1–30.

(Ates et al., 2011) M. Ates, S. Ravet, A.M. Ahmat, and J. Fayolle, 2011. An Identity-Centric Internet: Identity in the Cloud, Identity as a Service and Other Delights. ARES 2011, pp. 555–560.

(Bertino and Takahashi, 2011) E. Bertino, and K. Takahashi, 2011. Identity Management: Concepts, Technologies, and Systems. Artech House.

(BMVIT, 2016) BMVIT (2016). "Automated – Connected – Mobile. Action Plan Automotive Driving – Executive Summary". Online available: https://www.bmvit.gv.at/en/service/publications/downloads/action_automated_driving_2016-2018.pdf

(BMVIT, 2019) BMVIT (2019). "Austrian Action Programme on Automated Mobility". Online available from: https://www.bmvit.gv.at/en/service/publications/downloads/action_automated_mobility_2019-2022_ua.pdf

(Cao and Yang, 2010) Y. Cao, and L. Yang, 2010. A survey of Identity Management technology. In Proceedings of the IEEE ICITIS 2010, pp. 287– 293. IEEE.

(Cloud Security Alliance, 2011) Security Guidance for Critical Areas of Focus in Cloud Computing V3.0. CSA

(Creech and Hu, 2013) G. Creech and J. Hu. A Semantic Approach to Host-based Intrusion Detection Systems Using Contiguous and Discontiguous System Call Patterns. Computers, IEEE Transactions on, PP(99):11, 2013.

(Creech, 2014) Creech Gideon (2014). "Developing a high-accuracy cross platform Host-Based Intrusion Detection System capable of reliably detecting zero-day attacks" PhD Thesis. Engineering & Information Technology, UNSW Canberra. Online available: http://unsworks.unsw.edu.au/fapi/datastream/unsworks:11913/SOURCE02?view=true

(D'Ambros et al., 2010) D'Ambros, M., Lanza, M. and Robbes, R. (2010). "An extensive comparison of bug prediction approaches," In Proceedings of the 7th IEEE Working Conference on Mining Software Repositories, pp. 31–41.

(Dabrowski and Pacyna, 2008) M. Dabrowski, and P. Pacyna, 2008. Overview of Identity Management. Technical report, chinacommunications.cn.

(Damjanovic-Behrendt, 2018) Damjanovic-Behrendt, V. (2018). "A Digital Twin-based Privacy Enhancement Mechanism for the Automotive Industry", 2018 International Conference on Intelligent Systems (IS): Theory, Research and Innovation in Applications. Funchal, Madeira,

Portugal, 2018, pp. 272 – 279. DOI: 10.1109/IS.2018.8710526. Online available from: https://ieeexplore.ieee.org/document/8710526/

(Do et al., 2005) Do, H., Elbaum, S. and Rothermel, G., 2005. "Supporting controlled experimentation with testing techniques: An infrastructure and its potential impact," Empirical Software Engineering, vol. 10, no. 4, pp. 405–435, 2005.

(ENISA, 2016) ENISA (2016). Cyber Security and Resilience of Smart Cars: Good Practices and Recommendations.

(FPF, 2018) FPF (2018). Personal Data in Your Car. National Automobile Dealers Association and the Future of Privacy Forum. Online available: https://www.nada.org/personaldatainyourcar/

(Fromm and Hoepner, 2011) J. Fromm, and P. Hoepner, 2011. The New German eID Card. In Fumy, W. and Paeschke, M. (Eds.), Handbook of eID Security - Concepts, Practical Experiences, Technologies, pp. 154–166. Publicis Publishing, Erlangen.

(Green and Ateniese, 2007) M. Green and G. Ateniese, 2007. Identity-Based Proxy Re-encryption. In ACNS 2007, Vol. 4521 of LNCS, pp. 288–306. Springer.

(Haddouti and Kettan, 2015) S.E. Haddouti and M.D.E.-C. El Kettan, 2015. Towards an Interoperable Identity Management Framework: A Comparative Study. International Journal of Computer Science Issues (IJCSI), Vol. 12, Issue 6. ISSN: 1694 0814. Online: https://arxiv.org/ftp/arxiv/papers/1902/1902.11184.pdf

(Herraiz et al., 2009) Herraiz, I., Izquierdo-Cortazar, D. and Rivas-Hernandez, F. (2009). "Floss-metrics: Free/libre/open source software metrics," In Proceedings of the 13th IEEE European Conference on Software Maintenance and Reengineering (CSMR'09), pp. 281–284.

(Hitachi Systems Security, 2019) Hitachi Systems Security (2019). "Smart Car Security Threats: Is the Connected Car a Good Idea?". Online available: https://www.hitachi-systems-security.com/blog/smart-car-security-threats-is-the-connected-car-a-good-idea/

(Huang et al., 2010) H.Y. Huang, B. Wang, X.X. Liu, and J.M. Xu, 2010. Identity Federation Broker for Service Cloud. ICSS 2010, pp. 115–120.

(Jøsang and Pope, 2005) A. Jøsang and S. Pope, 2005. User centric identity management. AusCERT'05.

(Jus et al., 2014) Jus, R. Jalali, D. and Ernst, M.D. (2014). "Defects4j: A database of existing faults to enable controlled testing studies for java programs," In Defects4J: A database of existing faults to enable controlled testing studies for Java programs, San Jose: ACM, Jun. 2014, pp. 437–440, isbn: 978- 1-4503-2645-2. doi: http://dx.doi.org/10.1145/2610384.2628055. Online available from: http://defects4j.org.

(Kaler and McIntosh, 2009) C. Kaler and M. McIntosh, 2009. Web Services Federation Language (WSFederation) Version 1.2. OASIS Standard.

(Kamei et al., 2008) Kamei, Y. Monden, A. Morisaki, S. and Matsumoto, K.-i. (2008). "A hybrid faulty module prediction using association rule mining and logistic regression analysis," In Proceedings of the Second ACM-IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM '08), New York, NY, USA: ACM, pp. 279–281, doi: 10.1145/1414004.1414051. Online available from: http://doi.acm.org/10.1145/1414004.1414051

(Kariuki , 2019) Kariuki, D., 2019. "List of Blockchain-Based Self Sovereign ID Options" Blog post available from: http://www.cryptomorrow.com/2019/08/28/10-projects-in-blockchain-based-identity-management/

(Leitold et al., 2002) H. Leitold, A. Hollosi, and R. Posch, 2002. Security architecture of the Austrian citizen card concept. In ACSAC 2002, pp. 391–400.

(Levine, 2019) S.Levine, Satellite Finance, (2019).“Rising Worries about Connected Car Security Shifts M&A into Higher Gear”. Online available: https://www.satellitefinance.com/insights/rising-worries-about-connected-car-security-shifts-ma-higher-gear

(Lyons et al., 2018) Lyons T, Courcelas L, Timsit K (2018) “Blockchain for Government and Public Service”. (European Union Blockchain Observatory & Forum). Online available: https://www.eublockchainforum.eu/sites/default/files/reports/eu_observatory_blockchain_in_government_services_v1_2018-12-07.pdf?width=1024&height=800&iframe=true

(Lyons et al., 2019) Lyons T, Courcelas L, Timsit K (2019) “Blockchain and Digital Identity”. (European Union

Blockchain Observatory & Forum). Online available: https://www.eublockchainforum.eu/sites/default/files/report_identity_v0.9.4.pdf?width=1024&height=800&iframe=true

(Menzies et al., 2015) Menzies, T.J., Krishna, R. and Pryor, D., 2015. The promise repository of empirical software engineering data. North Carolina State University.

(Mockus, 2009) Mockus, A. (2009). “Amassing and indexing a large sample of version control systems: Towards the census of public source code history.,” In MSR, Vol. 9, pp. 11–20.

(Nabi, 2017). Nabi, A.G, 2017. “Comparative Study on Identity Management Methods Using Blockchain”. University of Zurich, Switzerland. Online available from: https://files.ifi.uzh.ch/CSG/staff/Rafati/ID%20Management%20using%20BC-Atif-VA.pdf

(NASA, 2004) NASA, 2004. Metrics data program data repository. Online available from: http://mdp.ivv.nasa.gov.

(NIST Blockchain, 2019) Loïc Lesavre, Priam Varin, Peter Mell, Michael Davidson, James Shook (2019). A Taxonomic Approach to Understanding Emerging Blockchain Identity Management Systems. NIST CYBERSECURITY WHITE PAPER (DRAFT) BLOCKCHAIN IDENTITY MANAGEMENT APPROACHES. Online available from: https://doi.org/10.6028/NIST.CSWP.07092019-draft

(Neuman et al., 2005) C. Neuman, T. Yu, S. Hartman, and K. Raeburn, 2005. The Kerberos Network Authentication Service (V5). RFC 4120 (Proposed Standard).

(Nunez and Agudo, 2014) D. Nunez and I. Agudo, 2014. Blindidm: A privacy-preserving approach for identity management as a service. International Journal of Information Security Vol. 13, Issue 2, pp. 199-215. Online: https://link.springer.com/article/10.1007/s10207-014-0230-4

(Nunez et al., 2012) D. Nunez, I. Agudo, and J. Lopez, 2012. Integrating OpenID with Proxy ReEncryption to enhance privacy in cloud-based identity services. In IEEE CloudCom 2012, pp. 241 – 248.

(Palfrey and Gasser, 2007) J. Palfrey and U. Gasser, 2007. CASE STUDY: Digital Identity Interoperability and e-Innovation. Berkman Publication Series.

(Sayyad and Menzies, 2005) Sayyad S.J. and Menzies, T.J. (2005) The PROMISE Repository of Software Engineering Databases. School of Information Technology and Engineering, University of Ottawa, Canada. Available: http://promise.site.uottawa.ca/SERepository

(Selvanathan et al., 2019) N. Selvanathan, D. Jayakody and V. Damjanovic-Behrendt (2019). “Federated Identity Management and Interoperability for Heterogeneous Cloud Platform Ecosystems”, In Proceedings of the 14th International Conference on Availability, Reliability and

Security (ARES'19), Workshop on Industrial Security and IoT (WISI 2019), August 26-29, 2019, Canterbury, UK. https://doi.org/10.1145/3339252.3341492

(Schmittner et al., 2019) Schmittner, C., Latzenhofer, M., Magdy, S.A., Bonitz, A., Hofer, M (2019). „Towards a Comprehensive Automotive Cybersecurity Reference Architecture". International Journal of Advances in Security, Vol. 12, No. 1&2.

(Śliwerski et al., 2005) Śliwerski, J., Zimmermann, T., and Zeller, A. (2005). When Do Changes Induce Fixes? In Proceedings of the 2005 International Workshop on Mining Software Repositories, Vol. 30. 1–5.

(Tavallaee et al., 2009) M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, "A Detailed Analysis of the KDD CUP 99 Data Set," Submitted to Second IEEE Symposium on Computational Intelligence for Security and Defence Applications (CISDA), 2009.

(Zwattendorfer et al., 2013) B. Zwattendorfer, K. Stranacher, and A. Tauber, 2013. Towards a Federated Identity as a Service Model. In Egovis 2013, pp. 43–57.

(Zwattendorfer et al., 2014) B. Zwattendorfer, T. Zefferer and K. Stranacher, 2014. An Overview of Cloud Identity Management-Models. In Proceedings of the 10th International Conf. on Web Information Systems and Technologies (WEBIST), pp. 82-92.

(Zwattendorfer, 2014) B. Zwattendorfer, 2014. Towards a Privacy-Preserving Federated Identity as a Service Model. PhD Thesis. The University of Graz, Austria.