IoT4CPS – Trustworthy IoT for CPS

FFG - ICT of the Future

Project No. 863129

Deliverable D5.5.1

Lifecycle Data Management Prototype I

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: Lifecycle Data Management Prototype
Type: Public
Editor(s): Violeta Damjanovic-Behrendt
E-mail: violeta.damjanovic@salzburgresearch.at
Author(s): Violeta Damjanovic-Behrendt (SRFG), Silvio Stern (X-NET), Nikolaus Dürk (X-NET), Christoph Schranz (SRFG)
Doc ID: D5.5.1

Amendment History

Version	Date	Author	Description/Comments
V0.1	12.03.2019	Violeta Damjanovic-Behrendt	Initial structure of the report
V0.2	01.04.2019	Violeta Damjanovic-Behrendt	State of the art and architecture described
V0.3	10.05.2019	Violeta Damjanovic-Behrendt	X-NET Figure 4 added and described in Section 5
V0.4	30.05.2019	Violeta Damjanovic-Behrendt	The report is ready for QA
V0.5	01.06.2019	Violeta Damjanovic-Behrendt	The report sent for QA
V0.6	05.06.2019	Mario Drobics	QA round 1
V1.0	30.06.2019	Violeta Damjanovic-Behrendt	The pre-final report sent for the second QA
V2.0	02.07.2019	Stefan Jaksic	QA round 2
V3.0	05.07.2019	Violeta Damjanovic-Behrendt	Final report

Legal Notices

The information in this document is subject to change without notice.
The Members of the IoT4CPS Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the IoT4CPS Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.

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.

Abbreviations

ADAS	Advanced Driver Assistance System
AI	Artificial Intelligence
AML	AutomationML (Machine Language)
API	Application Programming Interface
APM	Asset Performance Management
AR	Auto Regressive method
CAD	Computer Aided Design
CAEX	Computer Aided Engineering Exchange
CAM	Computer Aided Manufacturing
CIA	Confidentiality, Integrity, and Availability
CPPSs	Cyber Physical Production Systems
CPSs	Cyber Physical Systems
CRBM	Conditional Restricted Boltzmann Machine
CRM	Customer Relationship Management
DS	Dassault Systèmes
ERP	Enterprise Resource Planning
GDPR	General Data Protection Regulation
GE	General Electric
HMI	Human Machine Interfaces
IIoT	Industrial IoT
IoT	Internet of Things
JSON	Java Script Object Notation
KNN	K-Nearest Neighbour
MA	Moving Average method
MAEC	Malware Attribute Enumeration and Characterization
MES	Manufacturing Execution Systems
MISP	Malware Information Sharing Platform
ML	Machine Learning
MRO	Maintenance, Repair and Overhaul
OT	Operational Technology
PCA	Principal Component Analysis
PLCDM	Product Lifecycle Data Management
RDF	Resource Description Framework
SBI	Security-By-Isolation
SDN	Software-Defined Networking
SSNO	Semantic Sensor Network Ontology
STEP	STandard for the Exchange of Product data
STIX	Structured Threat Information eXpression
SVN	Support Vector Machine
VM	Virtual Machines
XML	eXtensible Markup Language
WoT	Web of Things

List of Figures

List of Tables

Executive Summary

The Automotive Industry is fragmented by different strategic alliances of manufacturers, which makes it difficult for the researchers to offer novel, alternative approaches to autonomous decision making, context- and situation-aware controls and self-adaptation, all of which are required to realise Smart CPSs for the Automated Mobility and the Automotive Manufacturing sectors. Hence, in IoT4CPS, we are designing a Digital Twin open source solution that will offer alternatives to proprietary technology stacks developed in corporate automotive research. To make the proposed solution accessible to a wider industrial and research audience, and to reduce the complexity of Digital Twin systems into smaller partitions of flexible, functionally independent and executable services, we follow the open source and microservices architectural styles.

This deliverable presents our methodology to design Digital Twins for security and safety validations. It is a summary of several recent IoT4CPS publications in which the authors address various aspects of the design of the Digital Twins for security and safety, including privacy aspects for Automated Mobility applications (for more details, see Appendix A "Publications by D5.5.1 Authors").

1. Introduction

This report presents a design of the first Digital Twin prototype to be developed in the IoT4CPS project (as part of Work Package WP5 “IoT Lifecycle Management”). The report begins with an overview of the most prominent definitions of Digital Twins, which are available at present in the literature, and of current design approaches addressing security and safety evaluations in the Automotive Industry. Furthermore, a general methodology that is specifically designed to address cybersecurity engineering, and cybersecurity and safety validations through the IoT4CPS Digital Twin prototype, is discussed in detail.

The first conceptual model of the IoT4CPS Digital Twin prototype and its adoption of the data acquisition method presented in the report D5.2 “Product Lifecycle Data Management (PLCDM) Stakeholder Perspective” are also discussed here. Based on the conceptual model of the Digital Twin prototype, we further design a microservices architecture of the Digital Twin prototype and a conceptual environment for testing the Digital Twin prototype as a virtual honeypot. The report closes with the steps to be performed for the coming D5.5.2 “Lifecycle Data Management Prototype II” document.

1.1 Motivation

Our focus in IoT4CPS WP5 is to design and implement a Digital Twin as a monitoring and supervision prototype for security and safety in the Automotive Industry, covering both the Automotive Driving and the Automotive Manufacturing (Industry 4.0) sectors. Specifically, our motivation is to design and implement an open source Digital Twin prototype to be interfaced with future European industrial platforms in the Automotive Industry sector.

The resulting Digital Twin prototype will be tested within complex industrial environments, incorporating sectorial governance models, security policies, GDPR (General Data Protection Regulation), and other guidelines, e.g. Article 29 Data Protection Working Party: WP29 Guidelines on Automated Individual Decision-Making and Profiling for the Purpose of Regulation 2016/679 (revised and adopted in February 2018). Through the design and development of such prototypes, we expect to deliver both scientifically relevant results and industry valid prototype solutions.

1.2 Relation to D2.2 Business Case on Security Verification Along the Life Cycle

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. The Device.CONNECT™ system 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 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. Hence, the IoT4CPS Digital Twin demonstrator is designed to serve the AVL’s business case (the Device.CONNECT™ system) for security, safety and privacy verification. In addition to cybersecurity perspectives the demonstrator addresses multi-stakeholder (multi-tenant) approach and the complexity of PLCDM stages of the real-world Automotive Driving applications to be “digitally twinned” in the project.

2. Literature Review

The Automotive Industry is heading towards a new paradigm that is designed to enable new technical features for safe, secure, comfortable and eventually, autonomous driving. Vehicles are becoming smart and responsive, equipped with traffic information services, infotainment features, driving assistance and danger-warning applications, safety controls, and the like. The Automotive Industry relies on technologies such as the Internet of Things (IoT), Web of Things (WoT), cloud computing, edge computing, fog computing, Big Data and Analytics, smart sensors, Cyber Physical Systems (CPSs), Digital Twins and Artificial Intelligence (AI) [1][2]. The role of IoT is to create and enable the collection of real-time sensor data that can be exchanged through the Web [3]. Edge and cloud computing technologies further support systems for advanced analysis and correlation of data. AI technologies enable data mining and the creation of added value through knowledge discovery. Big Data technologies provide systematic analysis of data generated along the entire product lifecycle and improve productivity of systems through rapid decision making [4][5]. CPS is another key technology that adds intelligence to traditional processes [6][7]. CPS integrates computational paradigms with the physical processes [8] and supports manufacturing capabilities such as reliability, interoperability, predictability and tracking. CPSs are expected to play a major role in the design and implementation of future software systems with new capabilities, as noted in [9][10]. The authors in [11] define a Cyber Physical Production System (CPPS) as an interconnected system operating "…across all levels of production, from processes through machines up to production and logistics networks". The term CPPS is often used as a synonym for the Smart Factory, emphasizing the scalable and modular structure of Smart Manufacturing applications [12].

Further advancements of CPSs are known as Smart CPSs, which are complex engineering systems that integrate heterogeneous hardware and software technologies through various analytics and decision-making mechanisms [13]. The concept of Smart CPSs is undergoing rapid changes to incorporate a plethora of challenges and desired features of the future automated systems, i.e. multi-tenancy, data sharing, data streams along lifecycle models, autonomous decision making, context- and situation-aware controls, self-adaptation, advanced cybersecurity, agile governance models for manufacturing platforms and their ecosystems, Digital Twins, and more.

2.1 Digital Twins Definitions

The evolution of microchip, sensor and WoT technologies opens the way for tracking smart products along their lifecycle phases, analysing the acquired data, extracting new knowledge from the corporate Web environments and communicating production and operating conditions [14]. Such technology evolution shifted the concept of Digital Twins from the aerospace industry to the Smart Manufacturing field [15]. The concept of twins in aircraft industries, has been used to support the optimization and validation technology of aircraft systems, based on the integration of sensor data, maintenance data and available historical/fleet data [16].

In 2002, Grieves coined the term Digital Twin that subsequently evolved from "conceptual ideal for PLM (Product Lifecycle Management)", "the mirrored space model", "the information mirroring model", to today's varying definitions. For example, Grieves in [17][18] defines a Digital Twin as "a set of virtual information constructs that fully describes a potential or actual physical, manufactured product from the micro atomic level to the macro geometrical level''. The authors in [19] define a Digital Twin as the use of holistic simulations to virtually mirror a physical system. A Digital Twin is also a "digital counterpart of a physical product" [15] and can be used for simulations and predictions of future states of the physical product, e.g. as discussed in [20]. The authors in [21] define a Digital Twin as a comprehensive digital representation of an individual product, its properties, conditions and behaviour.

The core functionality of a Digital Twin is to support design tasks and to validate system properties through multi-domain and multi-level simulations along lifecycle phases, including operational support [22]. The authors in [23] give a comprehensive overview of the Digital Twin definitions, which are currently available in the literature. As a virtual model of real industrial settings (including the Automotive Industry), the Digital Twin is expected to ensure information exchange throughout the entire lifecycle [15], enable virtualization of manufacturing systems [24], and automate decision making and system behaviour-based predictions [25].

2.2 Current Commercial and Open Source Digital Twin Solutions

To illustrate the current state of industrial practices, we additionally explore currently available designs and implementations of Digital Twins, which are built as either fully commercial or open source commercial solutions (i.e. Bosch IoT Suite that is based on Eclipse Ditto).

The most prominent examples of commercial software solutions implementing industrial Digital Twin technology are: General Electric, PTC Windchill, Dassault Systèmes, DXC, Siemens Simcenter, Microsoft Azure Digital Twins and Seebo Digital Twin.

- General Electric (GE) has developed the Digital Twin of a jet engine that enables the configuration of individual wind turbines, prior to their procurement and construction. Each virtual turbine is then fed with data from its physical equivalent. GE's Digital Twin is based on the Predix platform (www.predix.com) that delivers capabilities such as asset connectivity, edge technologies, analytics and Machine Learning (ML), Big Data processing, Asset Performance Management (APM), and implements asset-centric Digital Twins [26]. Apart Digital Twins, GE's Digital Twin Starter Kit is an open source toolset used to teach software engineers how to build Digital Twins. The Digital Twin Starter Kit is based on the Predix infrastructure and enables the design and hosting of Digital Twins. (GitHub page: https://github.com/DigitalTwin)
- PTC Windchill [27] has developed a smart PLCDM software called Windchill that supports failure reporting, analysis, and corrective actions.
- Dassault Systèmes (DS) has built an aerospace- and defence- manufacturing management software called "Build to Operate", which enables monitoring, controlling, and validation of all aspects of manufacturing operations [17].
- DXC has developed a Digital Twin for predicting performances of hybrid cars, before committing the changes in the car manufacturing process [28]. DXC has partnered with Microsoft to build ML solutions on an industrial scale, and used the Microsoft Cortana Intelligence Suite to run the Digital Twin, through continuous simulation of new ways for creating hybrid cars.
- The authors in [29] present the Simcenter 3D for Digital Twin. The Simcenter software distinguishes between (i) a product Digital Twin that allows users to virtually execute new designs and simulate the effects of the changes in the digital system, (ii) a production Digital Twin that is used to validate the effectiveness of a manufacturing process created for the factory floor assets, and (iii) a performance Digital Twin that processes Big Data from Industrial IoT (IIoT) products with the aim to improve product and production systems efficiency.

- Bosch IoT Things (https://www.bosch-iot-suite.com/) is a commercial product that uses the Digital Twin design methods for asset management, and for sharing device data and functionality across applications.
- Microsoft Azure Digital Twin (https://docs.microsoft.com/en-gb/azure/digital-twins/) is one of the recent Microsoft Azure cloud services used to create comprehensive models of the physical environment. In the manufacturing context, it can be used for predicting maintenance needs for a factory and for analysing real time energy requirements for an electrical grid.
- Seebo Digital Twin software (https://www.seebo.com/digital-twin-software/) provides a visual modelling tool for the graphical design of digital replicas of production line processes and assets. A functional Digital Twin prototype can be generated directly from the designed Digital Twin model. By using the Seebo IoT Simulator (https://www.seebo.com/iot-simulation/), the use cases, data flows Human Machine Interfaces (HMI) and predictive quality systems on the manufacturing assets, can be validated.

The open source community is at the forefront of creating software and hardware solutions for the experimentations in Industry 4.0 and Smart Manufacturing. However, there are currently only a few open source Digital Twin solutions available, among them Eclipse Ditto and CPS Twinning.

- Eclipse Ditto is an open source software solution that enables the design of Digital Twins in a form of IoT development patterns. Bosch IoT Things is an example of the commercial product that is based on open source Eclipse Ditto (GitHub page: https://github.com/eclipse/ditto).
- CPS Twinning is a framework for generating and executing Digital Twins that mirror CPSs, based on an AutomationML (AML) specification (GitHub page: https://github.com/sbaresearch/cps-twinning).

2.3 State-of-the-Art Implementations of Digital Twins for Cybersecurity and Safety Measures in the Automotive Industry

The concept of Digital Twins opens new paths to enhance security and safety of IoT and CPSs-based systems. The literature review on adopting Digital Twins for Information Security, shows several research directions with potential to overcome current security and safety challenges related to IoT and CPSs. For example, an approach presented in [30] compares configuration data of physical devices to their virtual replicas, in order to detect possible manipulated configurations and deviations that may refer to a compromised CPSs setting. The Digital Twin-based pen-tests could enable relevant tests virtually (instead of the real system), during both the operation phase and engineering phase with the aim to capture and fix vulnerabilities early [31][32]. Digital Twin concepts are also studied for privacy assessments and protection of the privacy of smart cars, e.g. in [33].

Finally, some tools that can be used for security research related to Smart CPSs and Digital Twins, include:

- The MiniCPS is a toolkit for security research on CPS networks that explores different attack scenarios and evaluate countermeasures [34].
- The CPS Twinning prototype presented in [30] contains security modules for network analysis and can be used to check certain security and safety rules in simulation mode.

3. Methodology for the Design of Digital Twins for Cybersecurity and Safety Engineering and Validations

The Digital Twin is a virtual counterpart to actual physical devices (assets) and their environments. It is instrumental in mimicking the asset's environmental conditions, processes, even diversity of interests of stakeholders ranging from certification bodies, test and governmental services, to insurance companies, and more. The desired capabilities of Digital Twins relate to monitoring of the PLCDM phases, prediction of the future states of assets, simulation of various conditions (including those conditions that are impractical to be created in real life), support for supply chain processes, financial decisions, etc. [35]). To achieve such capabilities, the Digital Twin requires various integration technologies and knowledge extraction methods to be orchestrated to support effective decision making, e.g. AI, real-time predictive analysis, and forecasting algorithms exploring Big Data derived from the IoT sensors and historical data. Through intelligent toolsets and methods, the Digital Twin can effectively optimise and improve the lifecycle processes and enhance asset performances, including their security and safety features.

The authors in [35] suggests the following elements to be considered for the design of Digital Twins:

- Asset modelling,
- Predictive analytics and decision-making methods, and
- Lifecycle-centric knowledge base with historical and real-time sensor data.

3.1 Asset Modelling

Asset modelling concerns architecting Digital Twins through designing the structure of assets (physical things) and their components, measurable parameters and production information about the assets (e.g. manufacturing date, maintenance history) [36]. Asset modelling adds value to connected sensor data and contributes to a range of new insights, e.g. obtaining information on asset health through sensors, which can be performed through inference, correlation and transformation of measured sensor values and asset states, conditions and maintenance records. It may also provide different presentation (visualization) forms for different user groups (stakeholders), e.g. one group of users may require the insight in operational data only, while others are more focused on individual devices. Adding information such as metadata, nearby environmental conditions, maintenance data, service history, configuration and production data, enterprise web services etc., contributes to a rich representation of the physical things and further augments the Digital Twin.

3.2 Predictive Analytics for Digital Twins

Predictive analytics comprise of a variety of techniques for calculating future outcomes based on historical and real-time data. It seeks to uncover patterns and capture relationships in data through techniques such as [37]:

- moving averages - discovers the historical patterns in the outcome variable(s) and extrapolate them to the future patterns, and
- linear regression - captures the interdependencies between outcome variable(s) and explanatory variables, and use them to create predictions.

Based on the underlying techniques, predictive analytics can be categorized into two groups [37]:

- Regression techniques (e.g., multinomial logit models) and
- ML techniques (e.g., neural networks, Supervised Learning, Unsupervised Learning, Reinforcement Learning [38]).

Predictive analytics techniques are primarily based on statistical methods, which often, when applied to massive data of Digital Twins, do not scale up in terms of computational efficiency. Big Data is characterized by factors such as heterogeneity, noise accumulation, spurious correlations, incidental endogeneity [39], and requires new statistical techniques to gain insights from predictive models. Some relevant approaches for cloud computing are based on task resource consumption patterns [40] and the usage of storage systems [41]. The analysis of behaviour patterns and derived models has been discussed in [42][43][44]. The authors in [45] present the Principal Component Analysis (PCA) technique used to retrieve relations between configuration and resource usage and performances in the cloud.

The core predictive models for behavioural analysis can be classified into two groups: location detection techniques and temporal behaviour analysis of time series. A variety of available location detection technologies leads to the massive accumulation of online data about users/assets location and their activity/usage histories. Such data can be used for mining knowledge in applications ranging from location-based recommendation systems to applications for tracking user/asset movements and activities. For example, pattern mining of GPS readings is often designed to identify specific patterns in a users' movement and behaviour [46]; the k-Means clustering algorithm is used to learn the user's significant locations and daily routines from the location history [47][48]; pattern mining from very large historical spatio-temporal dataset [49]; mining location patterns using Hidden Markov Models that can further feed frequent pattern mining methods, as presented in [50].

Regarding temporal behaviour analysis of time series, the most common approaches to modelling time series are: trend, seasonal, residual decomposition, frequency-based methods, Auto Regressive methods (AR), Moving Average (MA). In addition, the Conditional Restricted Boltzmann Machine (CRBM) represents a probabilistic model for time series used to solve a range of problems, from classification tasks to collaborative filtering and modelling of the motion capture [51][52].

3.3 Lifecycle-Centric Knowledge Base with Historical and Real Time Data

The Digital Twin's knowledge base needs to cover a wide range of diverse data: asset lifecycle data (e.g. location-based and time-series sensor data), data derived from analytics and decision-making algorithms, expert data, regulatory data, historical data. Such knowledge base is often augmented by adding data from a variety of third-party data sources, e.g. asset maintenance history from an Enterprise Resource Planning (ERP) system, account data from a Customer Relationship Management (CRM) system, environmental data, etc. One of the critical prerequisites to the knowledge base creation to support Digital Twins, is to have a proper data integration platform and infrastructure in place, enabling the integration of multiple data streams through standards and frameworks [35].

According to the size of a Digital Twin knowledge base and its maturity level, the authors in [36] differentiate among:

- a partial Digital Twin, with a small number of data sources that can be combined to infer further data (derivative data),
- a clone Digital Twin, with a larger amount of meaningful and measurable data sources and
- an augmented Digital Twin, that enhances connected asset data with derivative data and correlated data obtained from analytics tools.

A partial Digital Twin is a set of simplistic device models that could be implemented as JSON (Java Script Object Notation) documents with a set of observed and reported attributes (e.g. speed of a machine) and a set of desired values (e.g. an application is setting the speed of a machine) that can be correlated to detect operational abnormalities and instantly generate alerts. A clone Digital Twin is typically what is needed in industry: it is built on top of the product design and manufacturing information, and reflects its physical properties and uses real-time data [36].

3.4 Design Methodology for Digital Twins in IoT4CPS

To design Digital Twins to support security and safety engineering and validations in IoT4CPS, we suggest the following steps to be considered (see Figure 1).

STEP 1: Identification of assets and relevant security and safety objectives, and their modelling

This step is about asset identification and asset modelling through the design of the structure of assets (physical things) and components, measurable physical parameters and other digital features describing the assets (e.g. manufacturing date, maintenance history), including a different presentation forms for different stakeholders. The outputs of step 1 are asset data sets, asset lifecycle data (e.g. time-series sensor data), inferred and historical data related to assets.

This step is also about the identification and modelling of security and safety objectives of the system. This is an iterative process, driven by the contextual technical requirements of usage scenarios and application. The output are measures related to data confidentiality, integrity, and availability (CIA) for the core functional requirement of applications, as well as data related to role-based permissions and policies for the usage of applications.

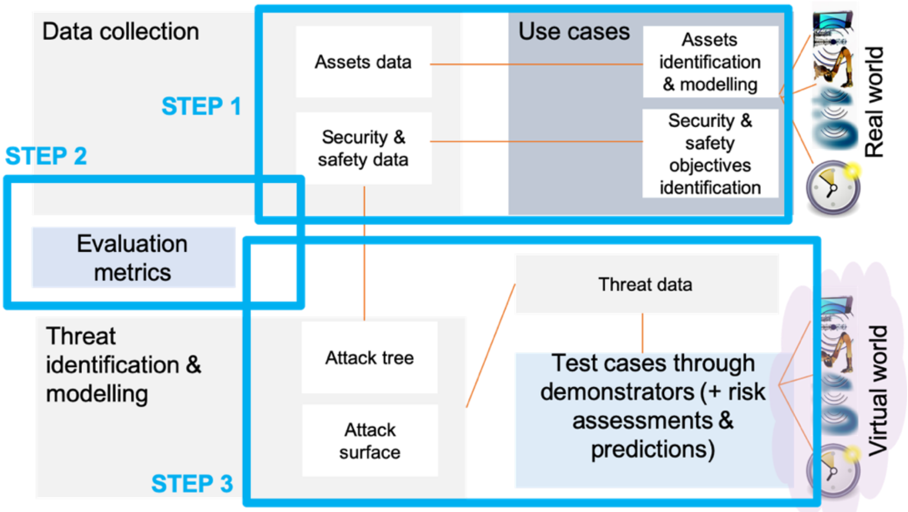

Data collection
STEP 1
STEP 2
Assets data
Security & safety data
Use cases
Assets identification & modelling
Security & safety objectives identification
Real world
Evaluation metrics
Threat data
Threat identification & modelling
Attack tree
Attack surface
Test cases through demonstrators (+ risk assessments & predictions)
Virtual world
STEP 3

Figure 1 - Design methodology for automotive security and safety validations through Digital Twins

STEP 2: Designing relevant security and safety evaluation metrics

Our method for security and safety validation is based on the multi-metrics security approach [15], according to which security metrics must be:

- measurable (consistently measured, with objective criteria),
- context specific, e.g. relevant to the CPS assets,
- expressed as a cardinal number, average or percentage, and (iv) expressed using at least one unit of measure.

The adoption of the multi-metrics approach suggests the further steps, which correspond to the steps of the proposed methodology for the design of Digital Twins in IoT4CPS, as illustrated in Figure 1.

- Step 1: Identify assets and assign them to the corresponding layers (e.g. network) and security indicators (e.g. CIA properties). This corresponds to the step 1 in Figure 1.
- Step 2: Select the right metrics according to security requirements and risks.
- Step 3: Analyse the metrics domain values (define value ranges). The steps 2 and 3 of the multi-metrics approach correspond to step 2 in Figure 1.
- Step 4: Risk assessment (the probability of occurrence of an undesirable event and the magnitude of its consequences over a specified period). This step of the multi-metrics approach corresponds to the step 3 in Figure 1.

STEP 3: Threat modelling and test case demonstrators based on security and safety risk assessment and forecasting

This step is about threat modelling and uncovering various cybersecurity and safety measures that serve as basis for the analytics methods in Digital Twin applications. In general, the analytics for Digital Twins consists of a predictive and a descriptive analysis of assets. Predictive analytics comprises a training phase (learning a model from training data) and a predicting phase (using the model for predicting future outcomes). The most used predictive models in ML belong to the category of Supervised Learning and include classification models for the evaluation of a discrete value (e.g. Logistic regression, Neural networks, Support Vector Machine (SVN)) and regression models for the evaluation of a numeric value (e.g. Linear regression model, Bayesian network and Naïve Bayes, K-Nearest Neighbour (KNN)) [53].

The functionality of a Digital Twin improves over time as more data get accumulated and processed through the effective ML algorithms.

3.5 Framing an Automated Mobility Case Using the Proposed Digital Twin Design Methodology

According to the Step 1 of the proposed methodology, all mission critical assets and applications need to be firstly identified (see Table 1).

Table 1: Assets in the Automotive Driving Case

Sensitive asset	Asset functionality	Security, privacy and safety concerns and mitigations
Infotainment controls	Navigation services and maps, entertainment services (audio/video), geo-fencing, cameras, traffic information, external media etc.	Revealing information about user's current location and navigation history, call history, geo-fencing data related to driving and working routines, heart rate and pulse, health data, banking accounts etc
Body Controls	Door/window locking, seat belt, seat heating	Revealing information about user's driving patterns and preferences, information about functionality of assets leading towards safety issues
Chassis controls	Driver alerts via Advanced Driver Assistance System (ADAS)	Revealing sensitive information linked to GPS data and traffic warnings, connected smartphone data, blind spots, audio alerts etc

In Table 2, the specific security and safety indicators need to be assigned to the assets, and the risk assessment for each mission critical asset and application need to be performed to select the right metrics. In addition, CPS risks and threats need to be evaluated based on metrics and models for cybersecurity defence, including the probability of their occurrence and the magnitude of possible consequences (see Table 2).

Table 2: Multi-Metrics Approach for the Automotive Driving Case

STEP 1	Asset identifications and their assignment to the correspondent layers and security indicators	
Assets	Security indicators	
Car sensor data	Data integrity must be guaranteed	
Statistical data	Data confidentiality must be protected	
Optimization configuration	Data integrity must be protected	
Central repository	Data confidentiality and integrity must be protected	
Optimization database	Data confidentiality and integrity must be protected	
HVAC controller database	Data confidentiality and integrity must be protected	
Historic database	Data confidentiality and integrity must be protected	
Expert knowledge database	Data confidentiality and integrity must be protected	
Sensors	Availability & integrity of sensors must be protected	
Actuators	Availability & integrity of actuators must be protected	
STEP 2	Select the right metrics according to security requirements & potential risks	
STEP 3	Analyse the metrics domain values	
Metrics	Description	Domain values
Digital Signatures (DS)	The ability to verify (DS) even when trusted 3rd party is not available	Min. key size for security is 3,072 bits
Data integrity in transit	The ratio of the read messages with assured integrity out of the total number of the read messages	Should be kept between 99-100%
The packet send ratio	The total number of transmitted packets	Should be kept between 95-100%
The packet delivery ratio	The total number of received packets	Should be kept between 95-100%
Round trip time	The time for a message to be sent & to receive an acknowledgement of that a message has been received	Between 100 and 200 ms
STEP 4	CPS risk assessment (the probability of occurrence of an undesirable event plus the magnitude of its consequences)	
Threat description	% of occurrence	Magnitude
(IoT4CPS) Attack on external devices connected to a vehicle (e.g. cell phone)	15 %	120

(IoT4CPS) Unintended transfer of data (information leakage)	7 %	-
(IoT4CPS) Extract Data/Code-unauthorized access to privacy information	3%	-

4. IoT4CPS Digital Twin Architecture

To enable Digital Twins to perform accurate simulations, analyses and prediction of events and situations related to an Automotive Driving ecosystem, the Digital Twin technology needs to fulfil the following requirements:

- Firstly, the Digital Twin technology requires a variety of data to be collected, analysed and sent back from a virtual space to the automotive ecosystem, in the form of decisions and information feedbacks.
- Secondly, the Digital Twin technology requires a collection of models to be created to describe product lifecycle phases (see e.g. [54]). The Digital Twin models vary from complex models, to simplified ones [22] and exist either as computational models (e.g. statistical packages for ML, analytics, optimization) or representational models (e.g. semantic data models, NoSQL, relational data models, relational derivatives, rule engines).
- Thirdly, the Digital Twin technology needs to be equipped with a collection of services to effectively monitor and simulate the physical world and to perform computations leading to decision making and knowledge discovery.

To reduce the complexity of the Digital Twin technology, we design the conceptual model along the above three design rationales [55]. As illustrated in Figure 2, the Virtualization Manager is at the heart of a Digital Twin and comprises of three building blocks: the Data Manager, the Models Manager, and the Services Manager. Apart from the Virtualization Manager, the Digital Twin contains another four management components: the Monitoring Manager, the Decision-Making Manager, the Simulation Manager, and the Interoperation Manager. The Monitoring Manager provides the connectivity between the Virtualization Manager and the assets of the factory shop floor. The Decision-Making Manager is responsible for the presentation formats of feedbacks created through analytics services of the Virtualization Manager. The Simulation Manager provides the simulation formats based on visualization dashboard services of the Virtualization Manager.

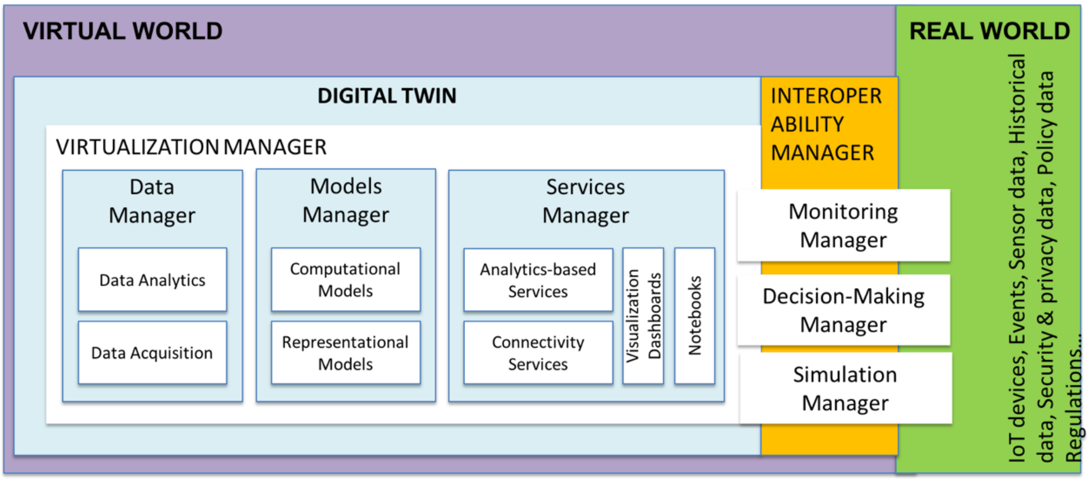


Figure 2 – Conceptual model of the Digital Twin prototype in IoT4CPS

The rest of this section descries the three building blocks of the Digital Twin's Virtualization Manager [55].

4.1 Data Manager

The Data Manager of the Digital Twin prototype contains the Data Acquisition and the Data Analytics components (see Figure 2).

- Data Acquisition: In Digital Twins, the collected data often comprise of real-time (or near real-time) sensor data, expert knowledge data, historical data and inferred data that are generated along the entire product lifecycle and aggregated in Big Data sets, data integrated from other enterprise systems and third-party systems. For example, the data collected from the design lifecycle phase in Smart Manufacturing include data for model building, model function, model design, Computer Aided Design (CAD), configuration & parameter optimization, structure, mechanics, size, material, history, predictions, simulations, processes, environment, faults, redesign activities, customers review and feedback (Tao et al. 2017, 2018). The data collected from the manufacturing lifecycle encompass manufacturing instructions, casting and moulding data, Computer Aided Manufacturing (CAM) planning data, and more.
- Data Analytics: The collected data of Digital Twins can be either structured, unstructured, or semi-structured. The data can be ingested as a stream in real-time, or as batch-oriented data generated from various sources. The data are often heterogeneous. In Digital Twins, the bigger the diversity of the collected data that the ML model has to analyse and learn about the states that matter along the manufacturing path, the better the model will be. For example, the availability of historical data helps ML models to learn the maintenance states of assets for predictive maintenance. In addition, the collected data could be used to predict product and process related behaviour, optimize manufacturing processes, discover anomalies, perform MRO (Maintenance, Repair and Overhaul) processes, etc.

The process of storing and preparing data for processing requires adequate analytics tools to be put in place, e.g., data can come to the Digital Twin system from Hadoop data clusters, SQL data exports, Kafka messaging server or other data stream processing engines. Data storage components and data formats (e.g. SQL, NoSQL, a data warehouse) can also have a profound impact on the capacity, performance, long-term reliability and durability of the Digital Twin data storage infrastructure. Finally, data failover and quality check mechanisms, backup and disaster recovery mechanisms need to be put in place and linked through Data Management capabilities of the Digital Twin.

4.2 Models Manager

The Model Manager of a Digital Twin includes data representation models (static, structural models) and data computation models (dynamic, behaviour models) (see Figure 1).

- Data representation models are used for storing, exchanging and searching through the data. These type of models includes [56] (i) semantic data models, e.g. ontologies and taxonomies for sharing PLM knowledge (e.g. Young et al. 2007); (ii) XML (Extensible Markup Language)-based models for encoding documents in a format that is human- and machine-readable (e.g. Choi, Yoon and Noh 2010); (iii) the STEP model (STandard for the Exchange of Product data) to describe product lifecycle data (e.g. Pratt 2001); (iv) the Computer Aided Engineering Exchange (CAEX), a meta model for the storage and exchange of engineering data models (e.g. Lüder, Hundt and Keibel 2010). The list can be further extended to include emerging manufacturing data representational models (Von Euler-Chelpin 2008), e.g. PLM XML, an open format from Siemens for facilitating PLM; ASME B5.59-2 standard that addresses performance and capabilities of machine tools at any time in their lifecycle phases, e.g., during specification, after acceptance testing, or during operation; ISO 16739 (IFC) defines a common data model for building lifecycle support that can be applied to manufacturing facilities; IEC 62890 defines standards for lifecycle management for systems and products used in industrial process measurement, control and automation, and many more.

- Data computation models perform analytics and processing along the product lifecycle phases, supporting e.g. system models, functional models, 3D geometric models, manufacturing computation models, usage models [56][15]. In Digital Twins, the data computation models need to support continuous learning and improvement based on run time data gathered from the operating CPSs. Practically, the collected data should be used to further improve simulation quality and adapt Digital Twins to contextual changes occurring in the system. By employing model learning algorithms (e.g. Deep Learning for Neural Networks to learn anomalies of the system), the inferred data learned during the run time can be incorporated into the Digital Twin knowledge base and continuous learning and improvement features of the Digital Twin can be further experimented to support a range of stakeholders involved in planning and designing, modifying, optimizing, and verifying industrial factory settings and processes.

4.3 Services Manager

The Services Manager of a Digital Twin requires a scalable and modular infrastructure to enable intelligent composition and orchestration of services. Digital Twin services may vary significantly, depending on business models and use cases, desired system capabilities, the role of stakeholders interacting with the Digital Twin, etc. For example, different stakeholders like operators, engineers, manufacturers, suppliers, customers, maintainers could all be interested in exploiting the Digital Twin, but each one from a different perspective.

Examples of Digital Twin services related to the Automotive Manufacturing domains are:

- production services for real-time state monitoring of the physical product, its environment and processes; real-time data management and asset management; real-time user management and user operations; real-time product failure analysis and prediction (anomaly detection); real-time behaviour analysis that can help manufacturers to improve product and production performances, e.g. condition monitoring, real-time image processing, etc.;
- supply chain control services that need to serve multiple tenants simultaneously; services that predict supply chain performances, etc.;
- cybersecurity services application security application (authentication, authorization, etc.), maintaining awareness of the security and privacy conditions through continuous monitoring processes of a Digital Twin; and more.

4.4 Microservices Architecture of the IoT4CPS Digital Twin Prototype

At present, Digital Twin platforms are still built as closed systems, thus limiting the overall acceptance and advantages in the Automotive Industry. Hence, our motivation in IoT4CPS is to design a flexible, open source solution for Digital Twins and to make it accessible to a wider industrial and research audience. In parallel, in contrast to monolithic systems that are often built as a massive code base, the microservice architectural style enables a single application to be developed as a suite of relatively small, consistent, isolated and autonomous services, each performing a specific task [57]. Microservices can be developed and deployed independently by different teams. They are language agnostic. The authors in [58] provide an analysis of the survey on major motivations for migrating from monoliths to microservices. For example, in the survey, software maintenance has been reported as very important by all the participants. Scalability of microservices, delegation of team responsibilities to other teams, and the easy support for DevOps were also highly rated.

In case of Smart CPSs and Digital Twins as complex and nonlinear systems that require a variety of mechanisms to represent their static artefacts and dynamic capabilities, the adoption of the modular architecture of microservices allows the application complexity to be reduced and code to be better maintained. However, the research in Digital Twin has raised a number of new challenges e.g. developing computationally efficient algorithms for predicting system behaviour in real-time, edge data processing tools, dealing with uncertainty in the system, etc. Microservices as a design choice offer flexibility and potential to reduce the complexity of Digital Twin systems. At the same time, using microservices can possibly open security vulnerabilities and threaten the trustworthiness of services, which requires a good design balance between security and system performance [59].

Our approach in IoT4CPS follows the microservices architectural style when decomposing the service- and application-logic and reducing the complexity of Digital Twins into smaller partitions of flexible, functionally independent and executable services. The conceptual model of the proposed architecture is shown in Figure 3. It consists of the following building blocks [60]: the Virtualization Manager, Data Manager, Models Manager, Services Manager and Interoperability Component, each of them encompassing a set of defined microservices. The rest of this section describes each of the Digital Twin microservices building blocks (Figure 3) [55].

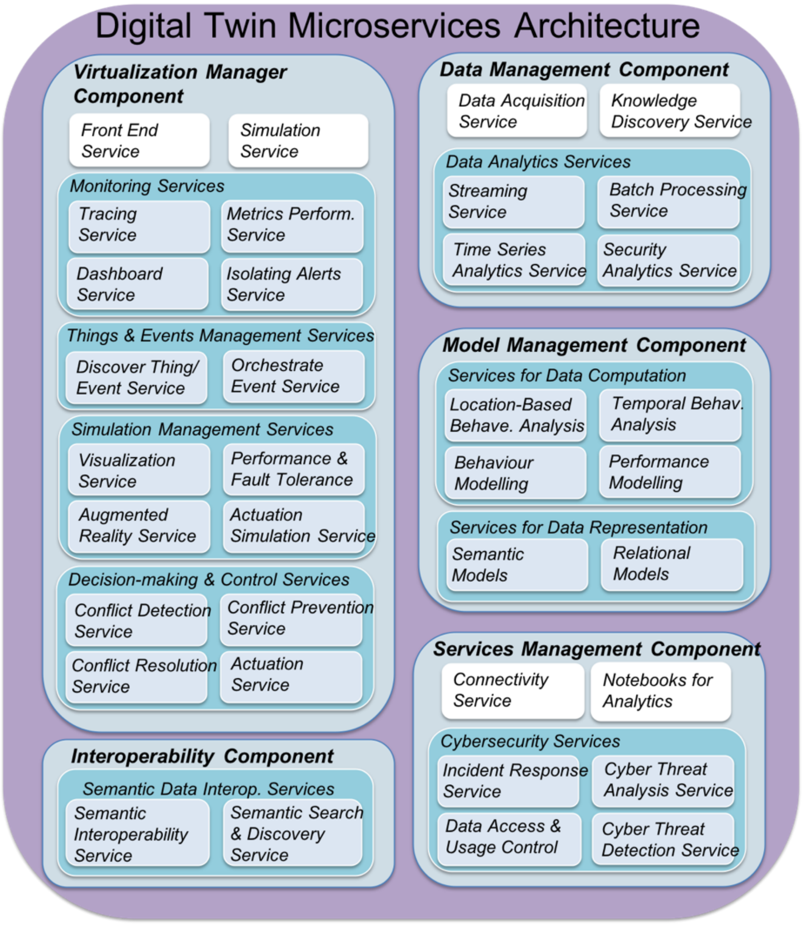

Digital Twin Microservices Architecture
Virtualization Manager Component
Front End Service
Simulation Service
Monitoring Services
Tracing Service
Metrics Perform. Service
Dashboard Service
Isolating Alerts Service
Things & Events Management Services
Discover Thing/ Event Service
Orchestrate Event Service
Simulation Management Services
Visualization Service
Performance & Fault Tolerance
Augmented Reality Service
Actuation Simulation Service
Decision-making & Control Services
Conflict Detection Service
Conflict Prevention Service
Conflict Resolution Service
Actuation Service
Interoperability Component
Semantic Data Interop. Services
Semantic Interoperability Service
Semantic Search & Discovery Service
Data Management Component
Data Acquisition Service
Knowledge Discovery Service
Data Analytics Services
Streaming Service
Batch Processing Service
Time Series Analytics Service
Security Analytics Service
Model Management Component
Services for Data Computation
Location-Based Behave. Analysis
Temporal Behav. Analysis
Behaviour Modelling
Performance Modelling
Services for Data Representation
Semantic Models
Relational Models
Services Management Component
Connectivity Service
Notebooks for Analytics
Cybersecurity Services
Incident Response Service
Cyber Threat Analysis Service
Data Access & Usage Control
Cyber Threat Detection Service

Figure 3 – Microservices Architecture of the Digital Twin Prototype in IoT4CPS

The Virtualization Manager

The Virtualization Manager is composed of the microservices shown in Table 3. It enables monitoring of assets and events through its Monitoring Manager. Through decision support services and controls, it detects conflicts and enables their resolution.

Table 3: Microservices of the Virtualisation Manager Component

Name	Description
Front-End Service	Establishes the front-end Web components of microservices. These are reusable components that can be imported into Web apps. The guidelines for creating web platform compatible Web components are given in [61].
Monitoring Services	A set of microservices that helps developers to understand the system behaviour by breaking the system down into smaller applications, e.g. Tracing Service, Metrics Performances Service, Isolating Alerts, Dashboards Service.
Tracing Service	Supports continuous tracking and tracing of factory shop floor assets across various subsystems, e.g. supply chains, ERP, Manufacturing Execution Systems (MESs), etc.
Metrics Performances Service	Measures performance and use of Operational Technology (OT) assets on the manufacturing shop floor.
Isolating Alerts Service	Provides controls for detecting and isolating problems that target specific processes/assets. It is coupled with Cybersecurity Services (of the Service Management Component) to create cross-layer alerts based on cybersecurity analytics.
Dashboards Service	Enables visualization dashboards for monitoring interactions, continuous integration, cross-service visibility, etc.
Things & Events Management Services	Enables discovery of things/assets and events, and the orchestration of events on the manufacturing shop floor.
Discover Things/Event Service	Allows for device functionality to be dynamically discovered and optimally exploited. It also supports the running events on the manufacturing shop floor to be discovered, for further services and decisions.
Orchestrate Event Service	Allows for creating more cooperative manufacturing models through effective orchestration.
Simulation Management Services	A set of microservices that incorporate performance measures and observations received from the physical world in order to manage simulation inputs for the DT.
Visualization Service	Enables visualization of measured performances of the system, e.g. through dashboards.
Augmented Reality Service	Enables simulations using Augmented Reality technologies.
Actuation Simulation Service	Enables simulations of a specific actuation.
Performance & Fault Tolerance Service	Measures performances and fault tolerance of the manufacturing system.
Simulation Services	Performs the simulation based on the formats defined by the Simulation Management Services.
Decision-Making & Control Services	A set of microservices for specific decision support and controlling functionalities of DTs.
Conflict Detection Service	Enables the identification of assets and events from the manufacturing shop floor, that significantly differ from the majority of relevant data (based on

	insights and measurements). Conflicts can be referred to as noise, or deviations.
Conflict Resolution Service	Based on identified conflicts, their nature, durability and other detected features, this service is in charge of providing the adequate resolution strategy.
Conflict Prevention Service	Through monitoring and analytics, this service provides mechanisms to avoid conflicts in the manufacturing system, caused by noise and deviation.
Actuation Service	Ensures that feedback created by DT mechanisms is transmitted to the real manufacturing environment.

The Data Management Component

The Data Management Component is composed of microservices dealing with data acquisition, data analytics and knowledge discovery, as shown in Table 4. Knowledge Discovery microservices require various analytics methods to be put in place. The results of analytics methods need to be further exposed to simulation and visualization microservices.

Table 4: Microservices of the Data Management Component

Name	Description
Data Acquisition Service	Enables data acquisition for DTs, e.g. data collected through sensors and from tracking and tracing technologies needs to be stored and maintained for warranty and other purposes.
Data Analytics Services	Enables various data analytics services, e.g. ML-based analytics for predicting assets' behaviour within the changed manufacturing environment.
Streaming Service	Supports streaming process analysis.
Batch Processing Service	Supports batch-oriented processing.
Time Series Analytics Service	Supports time series-based analysis.
Security Analytics Service	Supports security analytics and is further coupled with Cybersecurity Services of the Service Management Component.
Knowledge Discovery Service	The analytics techniques provide feedback mechanisms that send decisions and responsive actions back to the DT and physical system.

The Models Manager Component

The Models Manager Component includes services for the definition, execution and maintenance of data computation and data representation models, as shown in Table 5. Services for data computation are further coupled with Data Analytics Services. Services for data representation are maintained either as semantic models (e.g. described in RDF (Resource Description Framework)) or relational models (based on relational databases).

Table 5: Microservices of the Models Manager Component

Name	Description
Services for Data Computation	These services enable the major analytics processes of DTs.
Location-Based Behavioural Analysis	Typical services for location-based behavioural analysis allow for the factory floor assets to be identified based on their spatial location. The location of assets can be shared with other assets and events in the manufacturing ecosystem. These services include location prediction, location-based asset management, recommender systems, etc.

Temporal Behaviour Analysis Service	Ensures temporal localisation of the factory floor assets and events. Some popular methods include simulation techniques and discrete event systems (e.g. Petri nets).
Performance Modelling Service	Enables modelling of performances of the production line processes and assets.
Behavioural Modelling Service	Enables modelling of behaviour of the production line processes and assets under specific conditions of the manufacturing environment.
Services for Data Representation	These services support the inclusion of various data representation formats in DTs
Services for Semantic Models Management	Inclusion of the relevant manufacturing ontologies in the knowledge base, semantic services, semantic reasoning, ontology management for DTs, etc.
Services for Relational Models Management	Support of the management of relational data models; data interfaces and integration mechanisms for heterogeneous databases, etc.
Services for Data Computation	These services enable the major analytics processes of DTs.

The Services Management Component

The Services Management Component releases IoT/ WoT connectivity services, offers services through notebooks for customized analytics, and performs security controls related to data access and usage controls, threat detection service, threat analysis service, incident sharing and incident response service (see Table 6).

Table 6: Microservices of the Service Management Component

Name	Description
Connectivity Service	Ensures connectivity with factory floor assets, products and stakeholders involved through manufacturing life cycle processes. A comprehensive and systematic review of the literature is presented in [62], the most widely used standards and protocols for connectivity between machines in Smart Manufacturing are OPC, MQTT, IEC 62439, IEEE 1588, IEC 61850.
Notebooks for Analytics	Services that enable creating of notebook files that contain Python, Scala, R code and markdown text, which can be shared with other analytics services of DTs.
Cybersecurity Services	Services that enable cybersecurity analytics and management, including incident response, threat analysis, access controls and governance models. Enables different levels of security for the DT based on the sensitivity of data it carries. Smart CPSs are also vulnerable to security threats and privacy breaches that stem from (open source) communication technologies and protocols, and hence they require the right strategy to be followed to improve cybersecurity.
Incident Response Service	Enables different levels of security for the DT based on the sensitivity of data it carries. Smart CPSs are also vulnerable to security threats and privacy breaches that stem from (open source) communication technologies and protocols, and hence they require the right strategy to be followed to improve cybersecurity. Some examples of open source tools in this category are: • STIX (Structured Threat Information eXpression) language and serialization format for Cyber Threat Intelligence (CTI) (https://github.com/STIXProject) • MISP (Malware Information Sharing Platform) for sharing CTI (https://github.com/MISP) • MAEC (Malware Attribute Enumeration and Characterization) language for sharing information about malware (https://github.com/MAECProject/).

Cyber Threat Analysis Service	Cybersecurity threat analysis for a DT is based either on fully automated tools designed to quickly assess network traffic, file activity, etc. and contribute to the incident response process, or on static analysis of malware properties. An examples of open source tools in this category would be the Cuckoo Sandbox for automated malware analysis (https://cuckoosandbox.org/).
Data Access & Usage Control	In DTs, cloud service providers, users and often fog devices as tenants, do not trust each other. Hence, access and usage controls for data and services in DTs require well defined access control policies to preserve user privacy and ensure system security. Services in this category need to support Virtual Machines (VMs) e.g. providing an access control mechanism to avoid side-channel attacks, and to provide access controls for the fog and cloud, reciprocally [63].

The Interoperability Component

The Interoperability Component of the Digital Twin prototype is designed to offer interoperability mechanisms at the data level. In Digital Twins, interoperability services are critically important for enabling the usage of the various types of assets (and their data), that could additionally be produced by different manufacturers. Interoperability services ensure implementation of Digital Twins, and their simulation functionalities.

The authors in [64] address the following three levels of interoperability for WoT:

- semantic interoperability (decoding the meaning of data),
- structural interoperability (decoding the organization of data), and
- syntactic interoperability (converting data in a consistent way between a serialized representation and an internal data structure (e.g. a parse tree).

To enhance usability of data and data models in Digital Twins, the proposed Digital Twin prototype put its emphasis on semantic data interoperability, which is supported through the Semantic Interoperability Service, and the Semantic Search and Discovery Service (as shown in Table 7).

Table 7: Microservices of the Interoperability Component

Name	Description
Semantic Data Interoperability Service	This service is based on the W3C Web of Things (https://www.w3.org/WoT/) set of semantics and metadata standards around IoT. The focus of this service is on converting WoT representations that include identifiers, properties, and relationships into the meaning of data, through shared contexts, vocabularies, and ontologies (iot.schema.org, Semantic Sensor Network Ontology (SSNO) [65], SAREF, and more.)
Semantic Search & Discovery Service	In DTs, apart from ML-based processing of data, semantic search and reasoning capabilities are beneficial, too. This service ensures (i) rules and semantic alignments to transform data to the declared ontologies, e.g. using JSON-LD for RDF data serialization, and (ii) reasoning engines for inferring associations and links into the data [66].

5. Design Methodology for the IoT4CPS Digital Twin Demonstrator Realisation

To enable the use of the designed Digital Twin-based methods for security and safety controls in IoT4CPS, one of the important question to solve is about the experimentation environment to support virtual testing at multiple scales. Digital Twin prototypes in IoT4CPS is exposed to such an environment to collect security and safety related data and to learn about future network behaviour. This network environment includes the X-NET's Security-By-Isolation (SBI) gateway (SBI-GW in Figure 4), which enables further testing of Digital Twins as security honeypots. Practically, the SBI-GW will enable the enforcement of detected issues in data traffic, towards virtual Digital Twin-based honeypots that will directly analyse and learn attacker capabilities, in order to predict attackers' next steps and reconstruct the system back to the state before the attack. The attacks will be identified through implementation of analytical tools for anomaly detection and threat modelling.

The SBI gateway-based approach for testing the Digital Twin prototypes in IoT4CPS is in the design progress, and is shown in Figure 4.

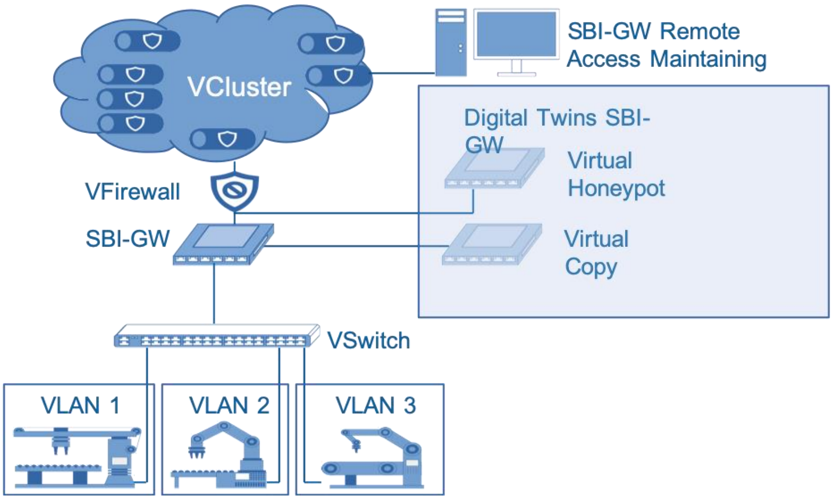


Figure 4 – SBI gateways for the realisation of Digital Twin prototypes as honeypots (X-NET)

6. IoT4CPS Digital Twin Demonstrator (Prototype Version 1.0)

The first prototype of the Digital Twin demonstrator in IoT4CPS is available from: https://git-service.ait.ac.at/im-IoT4CPS/WP5-lifecycle-mgmt
It is implemented as a simple client-based communications API (Application Programming Interface) called `demo_car_1` that "measures" temperature on the road (see Figure 5). The client connects to the GOST server (SensorThings Server) and the Apache Kafka server. As soon as the instances configured in the *instance.json* file are registered or updated in the GOST server, the temperature data is created.

```
config = {
          "client_name": "demo_car_1",
          "system": "at.srfg.iot-iot4cps-wp5.CarFleet1",
          "gost_servers": "localhost:8084",
          "kafka_bootstrap_servers": "localhost:9092"}
client = DigitalTwinClient(**config)

client.register(instance_file="instances.json"}
client.produce(quantity="temperature", result=23.4)

client.subscribe(subscription_file="subscriptions.json"}
received_quantities = client.consume(timeout=1.0)
```

Figure 5 – A simple client-based Digital Twin prototype version 1.0 (an excerpt)

To capture data from the external systems, e.g. from the other connected cars, the data needs to be distributed using streaming applications. In this way, `demo_car_1` that belongs to the `at.srfg.iot-iot4cps-wp5.CarFleet1` can acquire some selected data from an external `CarFleet2` and from the public weather stations (see Figure 9).

Figure 6 illustrates the flow of data between various stakeholders (car 1, car 2, weather station…) and data streaming applications. Practically, each tenant is connected to one streaming application that creates specific measurement data, and parses data sharing contracts into a distribution logic which is then deployed by the stream hub service. In the future, the data will include temporal and geospatial filtering criteria, in order to have more control over data flow to external tenants.

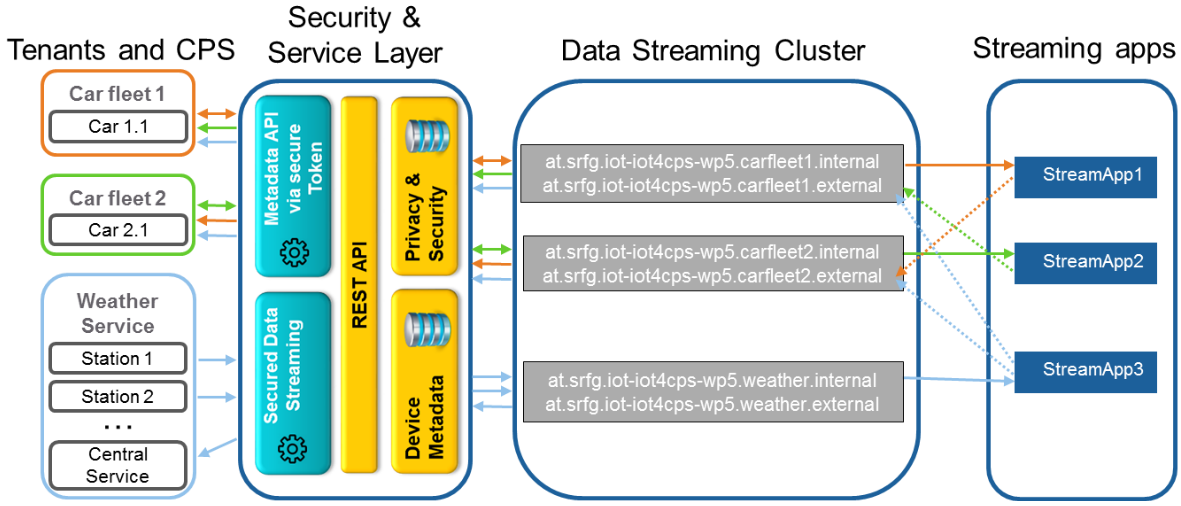
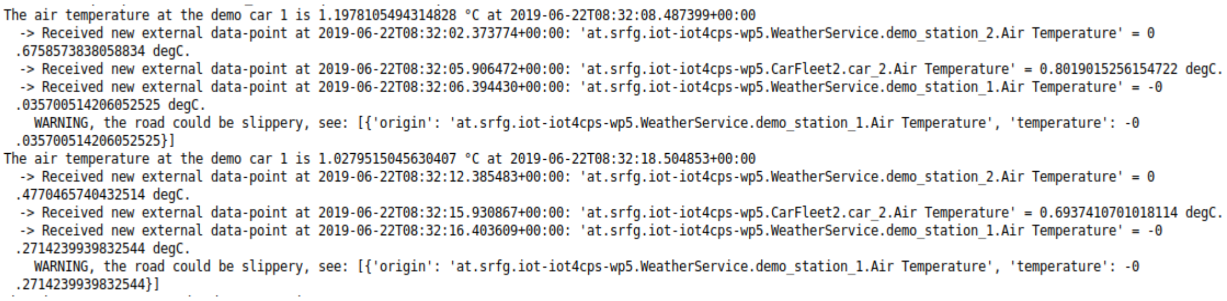

Tenants and CPS
Security & Service Layer
Data Streaming Cluster
Streaming apps
Car fleet 1
Car 1.1
Car fleet 2
Car 2.1
Weather Service
Station 1
Station 2
. . .
Central Service
Metadata API via secure Token
Secured Data Streaming
REST API
Privacy & Security
Device Metadata
at.srfg.iot-iot4cps-wp5.carfleet1.internal
at.srfg.iot-iot4cps-wp5.carfleet1.external
at.srfg.iot-iot4cps-wp5.carfleet2.internal
at.srfg.iot-iot4cps-wp5.carfleet2.external
at.srfg.iot-iot4cps-wp5.weather.internal
at.srfg.iot-iot4cps-wp5.weather.external
StreamApp1
StreamApp2
StreamApp3

Figure 6 – Multi-tenant stream processing in IoT4CPS

Figure 7 illustrates the outputs of the `demo_car_1` application that process data from other external tenants, e.g. `CarFleet2` and `WeatherService`.



```
The air temperature at the demo car 1 is 1.1978105494314828 °C at 2019-06-22T08:32:08.487399+00:00
  -> Received new external data-point at 2019-06-22T08:32:02.373774+00:00: 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_2.Air Temperature' = 0
 .6758573838058834 degC.
  -> Received new external data-point at 2019-06-22T08:32:05.906472+00:00: 'at.srfg.iot-iot4cps-wp5.CarFleet2.car_2.Air Temperature' = 0.8019015256154722 degC.
  -> Received new external data-point at 2019-06-22T08:32:06.394430+00:00: 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_1.Air Temperature' = -0
 .035700514206052525 degC.
    WARNING, the road could be slippery, see: [{'origin': 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_1.Air Temperature', 'temperature': -0
 .035700514206052525}]
The air temperature at the demo car 1 is 1.0279515045630407 °C at 2019-06-22T08:32:18.504853+00:00
  -> Received new external data-point at 2019-06-22T08:32:12.385483+00:00: 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_2.Air Temperature' = 0
 .4770465740432514 degC.
  -> Received new external data-point at 2019-06-22T08:32:15.930867+00:00: 'at.srfg.iot-iot4cps-wp5.CarFleet2.car_2.Air Temperature' = 0.6937410701018114 degC.
  -> Received new external data-point at 2019-06-22T08:32:16.403609+00:00: 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_1.Air Temperature' = -0
 .2714239939832544 degC.
    WARNING, the road could be slippery, see: [{'origin': 'at.srfg.iot-iot4cps-wp5.WeatherService.demo_station_1.Air Temperature', 'temperature': -0
 .2714239939832544}]
```

Figure 7 – The output of the "demo_car_1" application that processes the data from other tenants

7. Concluding Remarks and Next Steps

The adoption of Smart CPSs and Digital Twins is expected to significantly change traditional business models in the Automotive Industry. Although open source software and hardware technologies are becoming important for the collaborative design, development and maintenance of many industrial and engineering challenges, including those related to the integration of traditional Information Technology (IT) systems with Operational technology (OT) systems, there are still many computational challenges to be solved related to the complex systems based on Smart CPSs and Digital Twins.

The Digital Twin prototype presented in this report is still a research platform that is expected to evolve into an operational technology stack in the future. The prototype needs to be complemented with a comprehensive set of usage methods and validation metrics. The variability of metrics needs to be extended throughout lifecycle phases, addressing their specific security, privacy and safety measures. The role of ontologies in structuring the lifecycle-oriented knowledge base of Digital Twins needs to be explored, and performance and security metrics to be integrated into the knowledge base. The availability of large amounts of data needs to be tested for data quality and requires value-chain data governance mechanisms to be enforced, in order to identify the right amount of data to be processed. For example, some data sources can be periodically monitored, while others require to be traced in real time. Some data carry critical details for the system’s functionality, while others contain trivial details. More research is also needed to develop a proper strategy for maintaining the accuracy of Digital Twins, as their effectiveness in both the cloud computing and the fog computing is likely to be compromised otherwise.

References

[1] Wang, Shiyong, Wan, Jiafu, Li, Di, Zhang, Chunhua. 2016. “Implementing Smart Factory of Industrie 4.0: An Outlook.” International Journal of Distributed Sensor Networks, (4):1-10. doi: 10.1155/2016/3159805

[2] Cimini, Chiara, Pinto, Roberto, Cavalieri, Sergio. 2017. “The Business Transformation Towards Smart Manufacturing: A Literature Overview about Reference Models and Research Agenda.” In Proceedings of the International Federation of Automatic Control Conference (IFAC). 1452-1457. doi: https://doi.org/10.1016/j.ifacol.2017.08.2548

[3] Sarma, Sanjay, Brock, David L., and Ashton, Kevin. 2000. “The Networked Physical World.” Proposals for Engineering the Next Generation of Computing, Commerce & Automatic-Identification. White paper. Available: https://bit.ly/2sknSU5

[4] Davis, Jim, Edgar, Thomas, Porter, James, Bernaden, John, Sarli, Michael. 2012. “Smart Manufacturing, Manufacturing Intelligence and Demand-Dynamic Performance,” Computers & Chemical Engineering. 47, pp. 145–156.

[5] Lee, Jay, Kao, Hung-An, Yang, Shanhu. 2014. “Service Innovation and Smart Analytics for Industry 4.0 and Big Data Environment.” In Proceedings of CIRP, Elsevier 16, pp. 3–8, doi: https://doi.org/10.1016/j.procir.2014.02.001

[6] Lee, Jay, Bagheri, Behrad, Kao, Hung-An. 2015. „A Cyber-Physical Systems Architecture for Industry 4.0-based Manufacturing Systems,” Manufacturing Letters 3, 18–23. doi: https://doi.org/10.1016/j.mfglet.2014.12.001

[7] Jazdi, Nasser. 2014. “Cyber Physical Systems in the Context of Industry 4.0.” In Proceeding of the 2014 IEEE International Conference on Automation, Quality and Testing, Robotics. pp. 1–4. doi: 10.1109/AQTR.2014.6857843

[8] Lee, Edward A. 2008. “Cyber Physical Systems: Design Challenges.” In Proceedings of the 11th IEEE Intern. Symposium on Object and Component-Oriented Real-Time Distributed Computing, 363–369. doi: 10.1109/ISORC.2008.25

[9] Baheti, Radhakisan, and Gill, Helen. 2011. “Cyber-Physical Systems.” The Impact of Control Technology. 12, 161–166.

[10] Rajkumar, R., Lee, I., Sha, L., Stankovic, J. 2010. “Cyber-Physical Systems: The Next Computing Revolution.” In: Design Automation Conference (DAC), 47th ACM/IEEE. pp. 731–736. IEEE.

[11] Monostori, Laszlo. 2014. “Cyber-Physical Production Systems: Roots, Expectations and R&D Challenges.” In Proceedings of the 47th CIRP Conference on Manufacturing Systems, 17, 9–13.

[12] Weyer, Stephan, Meyer, Torben, Ohmer, Moritz, Gorecky, Dominic, Zühlke, Detlef. 2016. “Future Modeling and Simulation of CPS-based Factories: An Example from the Automotive Industry.” In Proceedings of the IFAC (International Federation of Automatic Control) Conference, 97-102.

[13] Horváth, Imre, and Gerritsen, Bart, 2012. “Cyber-Physical System: Concepts, Technologies and Manifestation,” In Proc. of the TMCE 2012, Vol. 1, 1–16.

[14] Schleich, Benjamin, Anwer, Nabil, Mathieu, Luc, Wartzack, Sandro. 2017. „Shaping the Digital Twin for Design and Production Engineering.” In CIRP Annals, Manufacturing Technol., 6(1), 141-144. doi: https://doi.org/10.1016/j.cirp.2017.04.040

[15] Rios, Jose, Hernandez, J.C., Oliva, Manuel, Mas, Fernando. 2015. “Product Avatar as Digital Counterpart of a Physical Individual Product: Literature Review and Implications in an Aircraft System.” In Proceedings of the 22nd International Conference on Concurrent Engineering

(ISPE CE2015), Vol. 2 of Advances in Transdisciplinary Engineering, 657–666. doi: 10.3233/978-1-61499-544-9-657

[16] Shafto, Mike, Conroy, Mike, Doyle, Rich, Glaessgen, Ed, Kemp, Chris, LeMoigne, Jacqueline, Wang, Lui. 2010. "DRAFT Modeling, Simulation, Information Technology & Processing Roadmap." Technology Area 11.

[17] Grieves, Michael. 2014. "Digital Twin: Manufacturing Excellence through Virtual Factory Replication." White paper. Available: https://bit.ly/2slbtiR

[18] Grieves, Michael, and Vickers, John. 2017. "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems." Kahlen F-J., Flumerfelt, S., and Alves, A., (Eds.) Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches. Springer International Pub., 85–113, doi: 10.1007/978-3-319-38756-7_4

[19] Rosen, Roland, von Wichert, Georg, Lo, George, Bettenhausen, Kurt D. 2015. „About the Importance of Autonomy and Digital Twins for the Future of Manufacturing." In Proceedings of the 15th IFAC Symposium on Information Control Problems in Manufacturing (INCOM 2015), 48(3):567–572.

[20] Gabor, Thomas, Belzner, Lenz, Kiermeier, Marie. 2016. "A Simulation-Based Architecture for Smart Cyber-Physical Systems." In Proceedings of the 2016 IEEE International Conference on Autonomic Computing (ICAC), 2016:374–379. Doi: 10.1109/ICAC.2016.29

[21] Haag, Sebastian, and Anderl, Reiner. 2018. "Digital Twin - Proof of Concept." Manufacturing Letters, 15(B), 64-66. doi: https://doi.org/10.1016/j.mfglet.2018.02.00

[22] Boschert, Stefan, and Rosen, Roland. 2016. "Digital Twin – The Simulation Aspect." In Hehenberger P, Bradley D, (Eds.) Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and their Designers, Springer International Pub., 9–74.

[23] Negri, Elisa, Fumagalli, Luca, and Macchi, Marco. 2017. "A Review of the Roles of Digital Twin in CPS-based Production Systems." In Proceedings of the 27th Int. Conf. on Flexible Automation and Intelligent Manufacturing (FAIM2017), Italy. Vol 11, 939-948.

[24] Schluse, Michael and Rossmann, Juergen. 2016. "From Simulation to Experimentable Digital Twins - Simulation-based Development and Operation of Complex Technical Systems." In Proceedings of the 2nd IEEE International Symposium on Systems Engineering (ISSE 2016), pp. 1-6, doi: 10.1109/SysEng.2016.7753162

[25] Kraft, Edward. 2016. "The Air Force Digital Thread/Digital Twin - Life Cycle Integration and Use of Computational and Experimental Knowledge." In Proc. of the 54th AIAA Aerospace Sci. Meeting, AIAA SciTech Forum. doi: 10.2514/6.2016-0897

[26] Predix. 2018. "PREDIX The Industrial IoT Application Platform." Technical Brief. Available: https://invent.ge/2LCgwmO (accessed 26 May 2018)

[27] PTC Windchill. 2018. "Windchill, From Requirements to Quality and from Manufacturing to Services." Available: https://bit.ly/2LM4BDd

[28] Overton, Jerry, and Brigham, Joan-Carol. 2017. "The Digital Twin. Data-driven simulations innovate the manufacturing process." White paper. Available: https://bit.ly/2xton2W (accessed 26 May 2018)

[29] Siemens Simcenter. 2017. "Predict Performance of Your 3D Geometry-based Design." Available: https://sie.ag/2xxfWUA

[30] Eckhart, M. and Ekelhart, A. (2018) Towards security-aware virtual environments for digital twins. In Proceedings of the 4th ACM Workshop on Cyber-Physical System Security (CPSS '18). ACM, New York, NY, USA, 61-72.

[31] Bécue *et al.,* (2018) CyberFactory#1 - Securing the industry 4.0 with cyber-ranges and digital twins. *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS)*, Imperia, 2018, pp. 1-4.

[32] Bitton R *et al*. (2018) Deriving a cost-effective digital twin of an ics to facilitate security evaluation. I Javier Lopez, Jianying Zhou, and Miguel Soriano (eds.), Computer Security, pp. 533-554, Cham, 2018. Springer International Publishing.

[33] Damjanovic-Behrendt V, (2018) A digital twin-based privacy enhancement mechanism for the automotive industry, In Proceedings of the 9th International Conference on Intelligent Systems: Theory, Research and Innovation in Applications. September 25-27, Madeira, Portugal.

[34] Antonioli, D. and Tippenhauer. N.O., MiniCPS: A Toolkit for Security Research on CPS Networks. In Proceedings of the First ACM Workshop on Cyber-Physical Systems-Security and PrivaCy (CPS-SPC '15). 2015. ACM, NY, 91–100. https://doi.org/10.1145/2808705.2808715

[35] Oracle. 2017. "Oracle: Digital Twin for IoT Applications: A Comprehensive Approach to Implementing IoT Digital Twin." Oracle White Paper. Available: https://bit.ly/2IZfFyY

[36] Kucera, Ryan, Aanenson, Mike, and Benson, Mark, 2017. "The Augmented Digital Twin: Combining Physical and Virtual Data to Unlock the Value of IoT." Exosite White paper. Available: https://bit.ly/2sirAws

[37] Gandomi, Amir and Haider, Murtaza. 2015. "Beyond the Hype: Big Data Concepts, Methods, and Analytics." Int. Journal of Information Management, 35(2), 137-144.

[38] Sutton, Richard, and Barto, Andrew. 2012. "Reinforcement Learning: An Introduction." The MIT Press, Cambridge.

[39] Fan, Jianqing, Han, Fang, and Liu, Han. 2014. "Challenges of Big Data Analysis." National Science Review, 1(2), 293–314.

[40] Mishra, Asit, Hellerstein, Joseph, Cirne, Walfredo, Das, Chita. 2010. "Towards Characterizing Cloud Backend Workloads: Insights from Google Compute Clusters." In SIGMETRICS Performance Evaluation Review, Vol. 37, 34-41.

[41] Aggarwal, Sonali, Phadke, Shashank, Bhandarkar, Milind. 2010. "Characterization of Hadoop Jobs Using Unsupervised Learning." In Proc. of the Cloud Computing Technology and Science (CloudCom), 748-753. doi: 10.1109/CloudCom.2010.20

[42] Bahga, Arshdeep, and Madisetti, Vijay Krishna. 2011. "Synthetic Workload Generation for Cloud Computing Applications." Journal of Software Engineering and Applications, Vol. 4, 396-410

[43] Chen, Yanpei, Ganapathi, Archana, Griffith, Rean, Katz, Randy H. 2010. "Analysis and Lessons from a Publicly Available Google Cluster Trace." In Proceedings of the EECS Department, University of California, Berkeley.

[44] Smith, James W. and Sommerville, Ian. 2011. "Workload Classification & Software Energy Measurement for Efficient Scheduling on Private Cloud Platforms." In Proceedings of the ACM SOCC 2011. Available: https://arxiv.org/pdf/1105.2584.pdf

[45] Yang, Hailong, Luan, Zhongzhi, Li, Wenjun, Qian, Depei. 2012. "MapReduce Workload Modeling with Statistical Approach." Journal on Grid Computing. 10, 2, 279-310. doi: 10.1007/s10723-011-9201-4

[46] Geng, Xiaoliang, Arimura, Hiroki, and Uno, Takeaki. 2012. "Pattern Mining from Trajectory GPS Data." In Proceedings of the 2012 IIAI International Conference on Advanced Applied Informatics, IIAIAAI 2012. 60-65.

[47] Ashbrook, Daniel, and Starner, Thad. 2002. "Learning Significant Locations and Predicting User Movement with GPS." In Proc. of the IEEE 6th Inter. Symposium on Wearable Computers (ISWC 2002), doi: 10.1109/ISWC.2002.1167224

[48] Ashbrook, Daniel, and Starner, Thad. 2003. "Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users." In Proceedings of the Personal and Ubiquitous Comp, 7:275–286.

[49] Tsoukatos, Ilias and Gunopulos, Dimitrias. 2001. "Efficient Mining of Spatiotemporal Patterns." In Proceedings of the International Symposium on Spatial and Temporal Databases (SSTD).

[50] Qiu, Weijun, and Bandara, Ayomi. 2015. "GPS Trace Mining for Discovering Behavior Patterns." In Proceedings of the 2015 International Conference on Intelligent Environments. IEEE Computer Society, USA, pp. 65-72. doi: 10.1109/IE.2015.17

[51] Mnih, Volodymyr, Larochelle, Hugo, Hinton, Geoffrey, E. 2011. "Conditional Restricted Boltzmann Machines for Structured Output Prediction." In Proceedings of the 27th Conference on Uncertainty in AI (UAI'11), F. Cozman and A. Pfeffer (Eds.). AUAI Press, 514-522.

[52] Taylor, Graham, W. and Hinton, Geoffrey, E. 2009. "Factored Conditional Restricted Boltzmann Machines for Modelling Motion Style." In Proc. of the 26th Annual International Conference on Machine Learning (ICML '09). ACM, 1025-1032.

[53] Bordawekar R., Blainey B, Puri R. (2015) Analysing analytics. Morgan & Claypool Publisher

[54] Kiritsis, Dimitris. 2011. "Closed-Loop PLM for Intelligent Products in the Era of the Internet of Things." Computer-Aided Design. 43(5), 479–501. doi: 10.1016/j.cad.2010.03.002

[55] V. Damjanovic-Behrendt, W. Behrendt, "An Open Source Approach to the Design and Implementation of Digital Twins for Smart Manufacturing". International Journal of Computer Integrated Manufacturing. Special Issue on Cyber Physical Systems with Applications in Production and Logistics. 2019. Taylor & Francis (impact factor 2017: 1.995). Published on Taylor & Francis. Online: https://www.tandfonline.com/doi/abs/10.1080/0951192X.2019.1599436

[56] Schroeder, Greyce, Steinmetz, Charles, Pereira, Carlos, Espindola, Danubia, 2016. "Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange." In International Federation of Automatic Control Symposium on Telematics Applications (TA 2016), 49(30), 12-17. doi: 10.1016/j.ifacol.2016.11.115

[57] Lewis, James, and Fowler, Martin. 2014. "Microservices." Available: https://martinfowler.com/articles/microservices.html (accessed: February 2019)

[58] Taibi, Davide, Lenarduzzi, Valentina, and Pahl, Claus. 2017. "Processes, Motivations, and Issues for Migrating to Microservices Architectures: An Empirical Investigation." IEEE Cloud Computing, 4(5), 22–32.doi:10.1109/mcc.2017.4250931

[59] Esposito, Christian, Castiglione, Aniello, and Choo, Kim-Kwang Raymond. 2016. "Challenges in Delivering Software in the Cloud as Microservices." IEEE Cloud Computing, 3(5).

[60] V. Damjanovic-Behrendt, "A Digital Twin Architecture for Security, Privacy and Safety". ERCIM News No. 115, Special Issue "Digital Twins". October 2018. Online: https://ercim-news.ercim.eu/en115/special/2103-a-digital-twin-architecture-for-security-privacy-and-safety

[61] W3C TAG. 2018. "Guidelines for Creating Web Platform Compatible Components." W3C Technical Architecture Group (TAG), Available: https://www.w3.org/2001/tag/doc/webcomponents-design-guidelines/ (accessed: February 2019)

[62] Liao Y., Deschamps F., Loures R. E., Ramos P. F. 2017. “Past, Present and Future of Industry 4.0 – A Systematic Literature Review and Research Agenda Proposal.” International Journal of Production Research, 55:12, 3609-3629.

[63] Zhang, Peng, Liu, Joseph K., Yu, F. Richard, Sookhak, Mehdi, Au, Man Ho, and Luo, Xiapu. 2018. “A Survey on Access Control in Fog Computing.” IEEE Communications Magazine, 56(2), 144–149.

[64] McCool, Michael. 2017. “The Web of Things and Semantic Interoperability.” IETF1000. Available at: https://bit.ly/2GisJ0M (accessed: February 2019).

[65] W3C SSNO, 2018: Extensions to Semantic Sensor Network Ontology. Online: https://www.w3.org/TR/vocab-ssn-ext/

[66] Szilagyi, Ioan and Wira, Patrice. 2016. “Ontologies and Semantic Web for the Internet of Things - A Survey.” IECON 2016 - 42nd Annual Conference of the IEEE Industrial Electronics Society.

Appendix A. Publications by D5.5.1 Authors

O. Veledar, V. Damjanovic-Behrendt, G. Macher, “Digital Twins for Dependability Improvement of Autonomous Driving”. In proceedings of the Workshop on Digitalisation of Industry, Infrastructure, and e-Mobility (co-located with the 26th European System, Software and Service Process Improvement & Innovation Conference (EuroSPI2019)), September 2019, Edinburgh, Scotland, UK (to be published)

V. Damjanovic-Behrendt, W. Behrendt, “An Open Source Approach to the Design and Implementation of Digital Twins for Smart Manufacturing”. International Journal of Computer Integrated Manufacturing. Special Issue on Cyber Physical Systems with Applications in Production and Logistics. 2019. Taylor & Francis (impact factor 2017: 1.995). Published on Taylor & Francis. Online: https://www.tandfonline.com/doi/abs/10.1080/0951192X.2019.1599436

V. Damjanovic-Behrendt, “A Digital Twin Architecture for Security, Privacy and Safety”. ERCIM News No. 115, Special Issue “Digital Twins”. October 2018.
Online: https://ercim-news.ercim.eu/en115/special/2103-a-digital-twin-architecture-for-security-privacy-and-safety

V. Damjanovic-Behrendt, “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, June 2018, pp. 272 – 279. DOI: 10.1109/IS.2018.8710526.
Online: https://ieeexplore.ieee.org/document/8710526/