A Digital Twin-based Privacy Enhancement Mechanism for the Automotive Industry

Violeta Damjanovic-Behrendt
Internet of Things
Salzburg Research
Salzburg, Austria
violeta.damjanovic@salzburgresearch.at

***Abstract*—This paper discusses a Digital Twin demonstrator for privacy enhancement in the automotive industry. Here, the Digital Twin demonstrator is presented as a method for the design and implementation of privacy enhancement mechanisms, and is used to detect privacy concerns and minimize breaches and associated risks to which smart car drivers can be exposed through connected infotainment applications and services. The Digital Twin-based privacy enhancement demonstrator is designed to simulate variety of conditions that can occur in the smart car ecosystem. We firstly identify the core stakeholders (actors) in the smart car ecosystem, their roles and exposure to privacy vulnerabilities and associated risks. Secondly, we identify assets that consume and generate sensitive privacy data in smart cars, their functionalities, and relevant privacy concerns and risks. Thirdly, we design an infrastructure for collecting (i) real-time sensor data from smart cars and their assets, and (ii) environmental data, road and traffic data, generated through operational driving lifecycle. In order to ensure compliance of the collected data with privacy policies and regulations, e.g. with GDPR requirements for enforcement of the data subject's rights, we design methods for the Digital Twin-based privacy enhancement demonstrator that are based on behavioural analytics informed by GDPR. We also perform data anonymization to minimize privacy risks and enable actions such as sending an automatic informed consent to the stakeholders.**

Keywords—Digital Twin, Machine Learning, Behavioral Analytics, CPS, Privacy Enhancement.

I. INTRODUCTION

The objective of this paper is to design the Digital Twin-based privacy enhancement methods and mechanisms for capturing privacy-related behaviour and anomalies acquired during smart car operational driving lifecycle, and for enabling informed consent, de-identification and anonymization, as strategies for improving subject's privacy rights and mitigating privacy vulnerabilities. The Digital Twin demonstrator presented in this paper, uses Machine Learning (ML) methods for behavioural analysis and forecasting of smart car operational processes and checks for compliance with the requirements of the General Data Protection Regulation (GDPR). In order to detect stakeholders, assets and privacy metrics of interest for the design of the Digital Twin demonstrator, we define a scenario that covers generic operational driving conditions, and from the designed scenario, we explain the use of ML-based feedback services targeting GDPR requirements on subject's rights.

The paper is organized as follows: Section 2 explores background technologies and the state of the art in the following areas: (i) core technology concepts of the Digital Twin, such as asset modelling, analytics, forecasting, decision making, and lifecycle knowledge base collections, (ii) ML methods for location- and temporal-behavioural analysis, and (iii) privacy methods related to Identity Management. Section 3 discusses the Digital Twin as a method for privacy assessment. Section 4 describes steps to achieve the Digital Twin-based privacy enhancing mechanisms. Finally, Section 5 concludes the paper.

II. BACKGROUND AND RELATED WORK

A. Digital Twin

The term Digital Twin was originally defined as "a virtual representation of what has been manufactured, through various lifecycle phases" [1]. The core benefits of the Digital Twin are summarized in the literature as follows [2]: it allows *visibility* in the manufacturing operations; it can be used to *predict the future state* of the machines; it can be used to *simulate* various conditions, which would be impractical to create in real life; it can be used to *connect with the backend business* applications to support supply chain, financial decisions, etc.

The Digital Twin is a virtual counterpart to actual physical devices (entities) that combines many Artificial Intelligence (AI)-based technologies and methods, real-time predictive analyses, and forecasting algorithms performing on top of Big Data derived from the Internet of Things (IoT) sensors and acquired historical data. The ultimate objective of the Digital Twin is to improve the design and execution in digital manufacturing through simulation, prediction of future states and intelligent decision-making related to various lifecycle phases. Through intelligent methods and toolsets, the Digital Twin enables manufacturing engineers to monitor the execution of simulated processes generated on top of multiple sensor data streams and to gain insights required to optimize the manufacturing processes, improve the entire lifecycle and enhance product performances on the fly.

The design of the Digital Twin suggests three major components to be considered [2]: (i) Asset modelling, (ii) Predictive analytics, forecasting and decision making, and (iii) Lifecycle knowledge base including both real-time sensor data and historical data.

978-1-5386-7097-2/18/$31.00

- Asset modelling is about architecting of the Digital Twin: designing the structure of its assets (physical things) and components, measurable physical parameters and other digital manufacturing information describing the assets (e.g. manufacturing date, maintenance history). Asset modelling adds value to connected sensor data and contributes to a range of new insights, e.g. obtaining an insight on health of sensors through inferring, correlating and transforming measured sensor values and asset states, conditions and maintenance records [3]. It may also include a different presentation (visualization) forms for different user groups, e.g. one group of users may require the insight about operational data, while the others could be more interested in individual devices.
- Analytics in Digital Twin applications 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)) [4].
- Lifecycle knowledge base of a Digital Twin collects asset lifecycle data (e.g. time-series sensor data), inferred and historical data. Practically, the functionality of a Digital Twin improves over time as more data is accumulated and processed by effective ML algorithms.

B. *Machine Learning*

ML has strong ties with IoT, enabling powerful analysis of multiple data streams collected from physical entities and used as a foundation for testing and prediction. The availability of historical data helps ML models to learn the maintenance states of assets for predictive maintenance. However, for continuous learning of the ML models, Digital Twins require a flow of real-time data. In the following, we overview location-based behavioural analysis and temporal behavioural analysis of time series, and ML methods for supporting behaviour and performance modelling.

- Location-Based Behavioural Analysis: The increasing popularity of location-based services is supported by a variety of location detection technologies, leading to the massive accumulation of online data about users' location and activity histories. Such data are used for mining knowledge in various applications, ranging from location-based recommendation systems to applications for tracking user's movements and habits [5-9], or pattern mining from very large historical spatiotemporal dataset [10-12], e.g. for methods created to discover pattern series of events [13].
- Temporal Behavioural Analysis of Time Series: The most common approaches to modelling time series include: (i) trend, seasonal, residual decomposition, (ii) frequency-based methods, (iii) autoregressive methods (AR), (iv) moving average (MA), (v) Box-Jenkins approach (ARMA model combining (iii) and (iv)). A recently proposed probabilistic model for time series, known as the Conditional Restricted Boltzmann Machine (CRBM), is used to solve a range of problems, from classification tasks to collaborative filtering and modelling of the motion capture [14-17]. Finally, the Multi-Label Learning and Classification model has been presented in [18].
- Behaviour and Performance Modelling Supported by ML: Three major categories of ML algorithms include (i) *Supervised Learning* (regression models and classification models), typically used to predict the behaviour of some system or individual, (ii) *Unsupervised Learning* (clustering and dimensionality reduction) used for observation grouping in order to form homogeneous clusters, and (iii) *Reinforcement Learning* models for mapping situations to actions to maximize some measures of utility for the system in question [19]. Reinforcement Learning (RL) has become one of the most active research areas in ML, AI and neural networks, with algorithms learning how to map situations to actions in order to maximize a numerical reward signal. In digital manufacturing, RL algorithms contribute to the extension of asset life and are used to lower maintenance and operating costs.

Methods that combine ML and data mining techniques have been exploited for modelling behaviour of distributed computing environments, e.g. a methodology that incorporates neural networks to construct and validate a nonlinear behaviour models have been presented in [20]. The authors in [21] use clustering to search for patterns in user-driven workloads in cloud computing. The analysis of behavioural patterns and deriving models for cloud computing systems has been addressed in [22-24]. Other relevant approaches for modelling behaviour and performances in cloud computing are focused on workload classification based on task resource consumption patterns [25] and the usage of storage systems [26]. Some other approaches on autonomic computing use ML for modelling system behaviour vs. hardware or software configuration [27].

C. *Privacy and Identity Management in Cloud Computing*

Privacy preserving approaches in cloud computing can be classified into two categories [28]: *cryptographic approaches* featuring encryption schemes and cryptographic primitives [29-32], and *noncryptographic approaches* with a policy-based authorization infrastructure [33]. The most common privacy approaches for verifying digital identities in cloud computing are presented in [34-35]. To overcome privacy issues with respect to cloud providers, the authors in [36-37] suggest to integrate proxy re-encryption with OpenId and SAML (Security Assertion markup Language). The authors in [38] describe privacy enhancing mechanisms for federated identity management systems. An example of a privacy-preserving identity and access management system for federated login that enables a controlled disclosure of the users' private information is the PRIMA system presented in [39-40].

III. Digital Twin for Privacy Assessment

Smart cars are designed to assist the drivers in a variety of ways, from improving their user experience (e.g. lane changing, parking assistance, night vision, traffic sign and traffic light recognition, map navigation support, etc.) to reducing distraction of drivers and improving their safety [41-42]. However, cloud-based technologies are exposed to a vast attack surface, partially through client applications and even more through various assets that could be vulnerable to attacks. In this section, we design a Digital Twin-based scenario that emphasizes the exchange of data related to assets and operation lifecycle processes in the Automotive Industry. The objective of our scenario is to define the exchange of privacy data to be processed by the privacy enhancement controls of the Digital Twin (see Figure 1), during the following phases:

- *initial phase*: A smart car collects various data including manufacturing data (information about the vehicle model, on-board sensors, tyre surface, etc.), driver's perception data (inertial cues, environmental cues), data about the connectivity to external systems and services (various stakeholders, e.g. retailer, insurance services, road services, etc.);
- *operational driving phase*: A smart car collects sensor data from various operational assets, environmental statistics data, authority reports (e.g. regulations);
- *analytics phase*: A Digital Twin of a smart car performs automated decision-making based on (near) real-time measurement and historical data;
- *reporting phase:* Currently, the decisions related to the improvement of operative driving procedures that affect user experience are under control of human engineers. The reports are often manually created and sent back to various stakeholders for further decisions. With the evolution of digital manufacturing, more intelligence and automation should be brought to business processes. In our scenario, the Digital Twin supports automatic creation of the decisions that are sent back to various stakeholders.

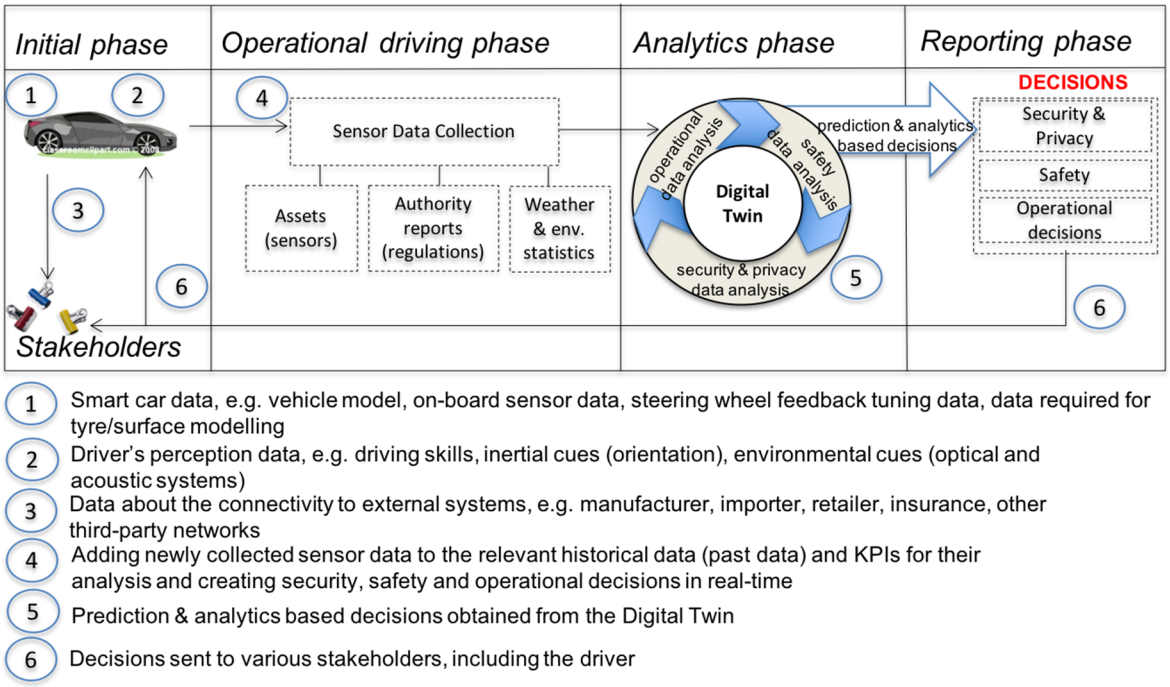

Fig. 1. Smart Car scenario with the Digital Twin.

The role of the Digital Twin, as shown in Figure 1, is to simulate and assess operational behaviour and performances of smart cars, replacing manually performed data analytics and reporting with automated decision-making support. Real time and historical data acquired from operational driving lifecycle, including external sources, environmental statistics and government regulations, are sent to the Digital Twin for analyses, e.g. location- and temporal behavioural analyses and their correlation through ML algorithms. The acquired data can contain privacy data that can affects the subject's (driver's) privacy in unexpected, unintended ways. Therefore, the aim of our scenario, is to demonstrate the Digital Twin as a privacy enhancement mechanism that correlates dynamic privacy data to various stakeholders' services, creates insights into privacy parameters and detects privacy anomalies to be automatically reported back to the subjects. The design and implementation of the Digital Twin toolset as an enforcement mechanism of the GDPR come with many requirements, from putting in place an effective data collection infrastructure, to establishing data models for supporting data management for stakeholders and assets, to designing privacy metrics and knowledge bases with implemented rules and policies, e.g. GDPR requirements.

In the following, we detect smart car stakeholders and assets of interest for the privacy assessment.

A. Detecting and Describing Stakeholders and Assets

Smart automotive systems are designed to augment user experience through information exchange amongst various stakeholders, which can open numerous privacy issues leading to reputational damage for car drivers, car manufacturers, suppliers, garages, network service providers, software and application providers, etc. In this section, we identify (i) stakeholders (actors) of a smart car ecosystem (see Table 1) and (ii) sensitive assets (e.g. the driver's smartphone) from the operational lifecycle (see Table 2).

Table 1 shows identified stakeholders, their roles, privacy concerns and risks associated to stakeholders' data.

TABLE 1. SMART CAR STAKEHOLDERS, THEIR ROLES AND RELATED PRIVACY CONCERNS AND RISKS (ANALYSIS PARTLY BASED ON [42]).

Stakeholders	**Stakeholders' Roles**	**Privacy Concerns & Risks**
Manufacturer	Provides the assembly of the car components.	Requires privacy-by-design approach; Privacy breach can be caused through the use of the aftermarket suppliers' components.
Supplier	Provides car components and/or operating system for connecting these components.	Requires privacy-by-design approach; Privacy breaches to be avoided by informing users how to correctly perform installations.
Aftermarket Supplier	Provides components with some additional features (e.g. media player).	Requires privacy-by-design approach with privacy controls of added apps in order to avoid privacy breaches.
Smart Car Users & Internal Services		
Driver	Drives and uses various smart car gadgets and applications/ services; Connects via his smartphone and other online gadgets; Actively uses external cloud applications.	Integrity breaches through apps for learning driver's habits; multimedia camera apps for sharing moments on the road; health cameras in the car; smart backup cameras for monitoring the roads, etc.
Passenger	Uses gadgets and apps or is exposed to them.	Integrity breaches through unintentional data sharing or leakage of data.
External Services		
Road services	Monitoring road and traffic conditions; Safety recommendations and contextual insights, e.g. speed limit changes, roadway conditions.	Access to privacy data; Mining and inferring through available privacy data; Predicting user behaviour based on privacy data.
Marketing services	Monitoring driving habits and preferences in order to create personalized offers.	Access to privacy data for various marketing analyses; Insights from historical data (including privacy data) to understand various trends.
Test & certif. services	Monitoring driving habits; Contextual insights.	Inferring privacy data; Forecasting based on privacy data.
Insurance services	Pay-How-You-Drive insurance plan.	Inferring privacy data and predicting user behaviour; Scheming users for socio-demographic informed offers.
Energy/fuel services	Energy/fuel supply.	Informational leakage and/or data sharing.
Network connectivity providers & services	Network access and services.	Loss of privacy information in the cloud.
Smart cities & services	Economical use of road infrastructure.	Use of historical and real-time data to create better user experience; Inferring privacy data.

Governmental bodies and regulations strongly encourage smart car manufacturers to make cybersecurity a priority. The following initiatives strongly influence the automotive sector:

- eCall system requires from all new cars sold in EU to be equipped with an eCall alarm system that, in case of an accident, can automatically call the emergency services and send the location of the car.
- GDPR regulatory requirements are related to privacy data protection in EU, and
- the NIS Directive (the Directive on Security of Network and Information Systems) affects search engines, cloud providers and online marketplaces, and set cybersecurity regulations, incident response procedures, etc.

The next step is about identification and description of assets from the smart car operational driving lifecycle. Table 2 summarizes identified assets with the ability to produce sensitive data, their core functionality and associated privacy concerns and risks to various stakeholders of the operational lifecycle.

TABLE 2. SMART CAR ASSETS, THEIR FUNCTIONALITIES AND RELATED PRIVACY CONCERNS AND RISKS (ANALYSIS PARTLY BASED ON [42]).

Assets	Asset Functionality	Related Privacy Concerns
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 belts, heating seats	Revealing information about user's driving patterns and preferences.
Chassis controls	Alerts sent to drivers via ADAS (Advanced Driver Assistance Systems)	Revealing sensitive information linked with GPS data and traffic warnings, connected smartphone data, blind spots, audio alerts, etc.
Power Train controls	Speed control	Driving patterns and preferences; speed.
Communications controls	Authentication features; Connectivity with external services through an embedded GSM module or driver's smart phone; Stolen vehicle tracking; emote engine start; etc.	Revealing privacy data stored on smartphones; Sharing privacy data via smartphone with: service providers, developers, criminals hacking physical access to device, behavioural marketing; government; geotags GPS capabilities to embed exact location into posts or photos, etc.
Smartphones	Authentication features; Connectivity with smart car services and applications using e.g. tethered connections, etc.	Unauthorized access and manipulation of information stored on smartphone, e.g. e-mails, instant messenger apps, location data and history, call history, online accounts and passwords typed into phone, photos…

B. *Privacy Scenario Design for Digital Twins in Operational Driving Lifecycle*

The majority of connected infotainment controls in smart cars collect various personal data, from data about driver behaviour and habits that can be used for the analysis and the adjustment of car's settings for drivers, to on-board sensor navigation systems informing about the precise location of the car, where it has been previously parked and for how long. The driving patterns that can be identified based on collected personal data help advertisers and insurance companies to set their plans with special offers and discounts and to place new offers in a way that is recommended through tracking systems. Such data can be further processed, combined and correlated to other sensitive information, e.g. bank accounts, health information, etc.

Our privacy scenario is designed to collect privacy data from the Advanced Driver Assistance System (ADAS) that interprets driving patterns, typical routes and habits of the drivers in order to support augmented customization of smart cars and better driving experience. We design the Digital Twin that combines and correlates privacy related data and infers privacy anomalies.

Figure 2 illustrates an interaction flow amongst smart car (with real-time sensor data measurements and collections), Digital Twin (with historical data and ML analytics models) and a formal knowledge base containing GDPR clauses to be processed in order to detect and assess privacy concerns and risks. In Figure 2, smart car (presneted as the real-world asset) and its digital representation (a Digital Twin) are connected via cloud and edge computing. Smart car is designed to be managed at the edge, capturing various alerts and states measured by the car's on-board sensors. In the cloud, the Digital Twin captures all the information referring to the operational driving state of the car, including its manufacturing and operational history, and sends this information to stakeholders.

The real-time insight into the operational driving states provided by the Digital Twin, enables operational efficiency and automated feedback to drivers. As edge and cloud computing raise security and privacy issues, the goal of designing our Digital Twin demonstrator is to detect and assess privacy data, enhance related privacy-by-design processes and provide GDPR compliance. The overall outcome should be quicker and more informed decisions on possible privacy vulnerabilities related to operational driving lifecycle.

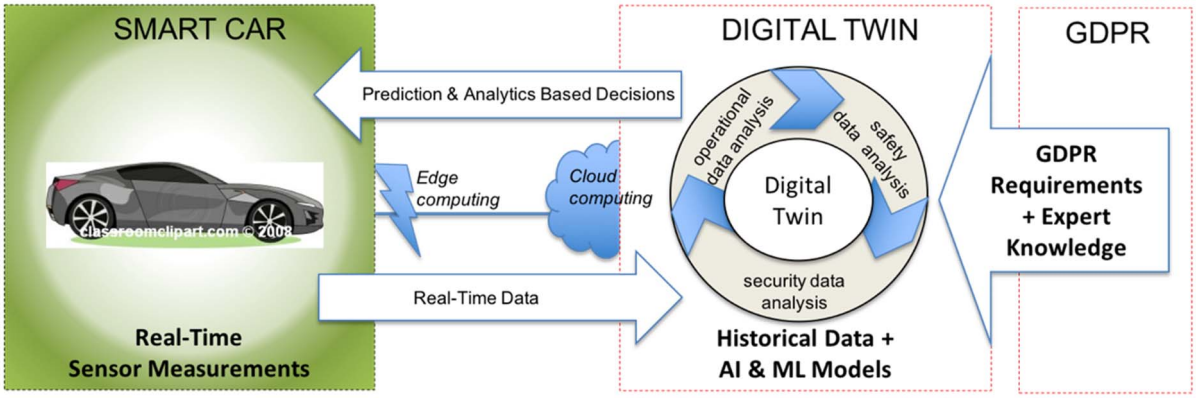

Fig. 2. Privacy scenario with the Digital Twin and supported by the GDPR requirements and expert knowledge (rules and policies).

Figure 3 gives more detailed view of privacy aspects of the automotive scenario presented in this paper. The smart car driver uses infotainment controls through an interactive connected cockpit and a smartphone. The core assets related to infotainment controls, their functionalities and the associated privacy concerns and risks are summarized in Table 2. For example, infotainment-based navigation assets and external media connected to the infotainment components (e.g. smartphones) could cause privacy breaches through loss of information in the cloud, thus damaging the car manufacturer's reputation and compromising driver's (and passenger') confidential and privacy information. Any action performed on a smartphone and information stored on the phone is under a risk of being manipulated and accessed in an unauthorized way. Privacy data captured via smartphone are accessible to various stakeholders, from service providers, behavioural marketing, geo-tagging services that embed location into posts and photos, insurance services, etc.

Figure 3 shows an insurance service (presented as stakeholder) that observes the data containing information on how the car is driven (e.g. speed and sudden acceleration and deceleration, eCall system history about road accidents, location history) and correlates these data to the car driver's smartphone data, with the possibility to access variety of privacy data (e.g. the address and the name of the car driver, his/her bank details). Based on inferred knowledge, the insurance services create a special "Pay How You Drive" insurance premium to be offered to the driver. Note the scenario illustrated in Figure 3 doesn't have the functionality of the Digital Twin to assess privacy related anomalies.

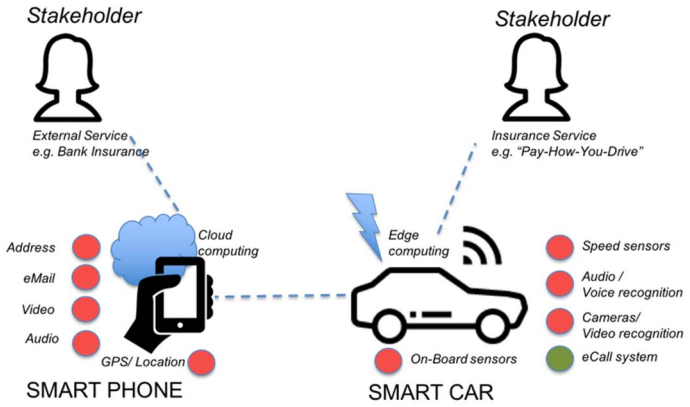

Fig. 3. Privacy Scenario (without a Digital Twin).

The recent introduction of the GDPR raised complex issues on subject's rights related to the control of personal data, e.g. subject's rights to access their personal data (Article 15), rights to rectify inaccurate personal data (Article 16) and erase personal data (Article 17). From the perspective of social media and external services such as online marketing and targeting, insurance services, etc., "*transparent information and communication with the data subject* (Article 12) [need to be provided] *when personal data are collected from the data subject* (Article 13) *or have not been obtained from the data subject* (Article 14)". In other words, to be GDPR compliant, external business services should check for permissions to use the subject's personal data and should provide a reason for it.

Figure 4 illustrates the same scenario previously shown in Figure 3, which now enables privacy assessments based on specifically designed functionality of the Digital Twin. Here, an insurance service observes data containing information on how the car is driven and before further combining these data with data obtained from the car driver's smartphone, the Digital Twin-based privacy enhancement mechanisms check for privacy concerns, to prevent and minimize possible risks of privacy breaches.

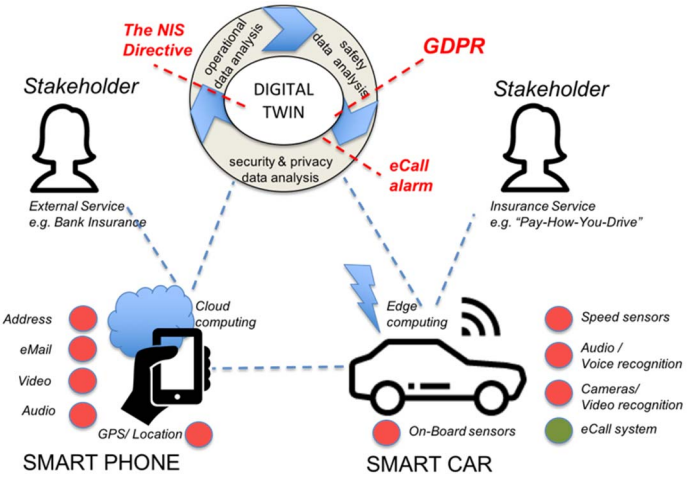

Fig. 4. Privacy Scenario with the Digital Twin privacy assessments providing GDPR compliance.

IV. Privacy Enhancing Mechansims for Digital Twins

To learn, simulate and predict privacy aspects of smart car's behaviour and performances, the ML-based privacy enhancement mechanisms of the Digital Twin demonstrator consider data models including (i) real-time sensor measurements (data) collected during the operational driving lifecycle, (ii) historical data related to operational lifecycle, and (iii) GDPR requirements implemented as an integrated expert knowledge base (see Figure 2). Assets that contains sensitive and privacy data need to be identified, while privacy anomalies need to be detected and assessed through an iterative process that uses ML algorithms to learn each of the car's assets over time. The supervised ML algorithms are trained to classify data into anomalous and nonanomalous data (e.g. k-Nearest Neighbours, Decision trees or Bayesian networks can be used to encode probabilistic relationships between the variables).

An alternative to the supervised ML is known as *Unsupervised ML* that features algorithms which are sensor/asset agnostic and create systems that are trained to detect sensitive data and privacy anomalies, without the need to learn the underlying monitoring process. In other words, the supervised ML-based predictions simulate operational models/ processes based on the car's historical data and the GDPR expert knowledge base. The unsupervised ML models enable an algorithm-driven learning of sensor data, without the need for historical data.

In case of *Reinforcement ML*, the detection and assessment of privacy anomalies require capturing of the most important aspects of the privacy problem that one agent (e.g. the Digital

Twin) learns through its interaction with the environment. In other words, the learning agent must be able to understand the nature and the state of privacy data and take an action that affects the state, e.g. to assess privacy data and send a request for informed consent.

The underlying privacy techniques for implementing GDPR requirements [43] are known as anonymization [44] and pseudonymization [45]. Our methodology for enabling GDPR compliance mechanisms for Digital Twins includes the following five (5) steps workflow:

- Step 1 is about identification of stakeholders and assets, including external services and knowledge bases. This step is followed by the collection of structured and unstructured data from multiple sources that will be made ready for analysis using various methods for data merging, reshaping, transforming, etc. Here, we also identify specific privacy metrics and perform the *data type identification* in order to detect context and privacy attributes, e.g. names, e-mails, addresses, personal codes, etc. We associate metrics with the data types;
- Step 2 is about *vulnerability detection* which is a process of correlating data columns to find out unique values that appear occasionally (e.g. a combination of the age, location, school and race). In this step we exploit ML algorithms that can be used for anomaly detection in time-series of various types, e.g. multivariate time-series from speed measurement sensors, and demands coming from external services and third parties, e.g. insurance services for *profiling* the subjects in order to optimize their decisions affecting the subjects (as defined in revised guidelines addressing profiling and automated decision-making under the GDPR, see [46]);
- Step 3 is about providing GDPR compliance through *de-identification of personal data* by replacing privacy data with fictional values. De-identification of personal data ensures compliance to GDPR requirements, but does not protect against privacy breaches;
- Step 4 is about interpreting the resulting data, which in our case requires performing actions like: *sending informed consent requests to data subjects*;
- Finally, in step 5, we use *data anonymization* to minimize the privacy risks and protect subjects from privacy breaches. For data anonymization we use syntactic approaches that operates under the k-anonymity principles, as suggested in [47-48].

V. Conclusion

The Digital Twin is a virtual representation of the real-world manufacturing, operational, logistic, maintenance, even administrative processes related to either a company or a product, that is designed with the aim to improve real-world products and processes based on simulated data and ML supported decisions. Designing the Digital Twin demonstrator for the automotive sector and smart cars augments overall functionality of vehicles, including their security and safety, while privacy data should be kept secure in the cloud and should comply with privacy policies, procedures and regulations such as the GDPR, the NIS Directive, and the eCall alarm system at the European level, as well as regulations at relevant national data protection laws.

In this paper, we designed mechanisms for detecting privacy anomalies in the automotive ecosystem and for minimizing further privacy risks. Such mechanism is designed as part of the Digital Twin demonstrator. Further work is about implementation of a Digital Twin prototype with the proposed privacy enhancing toolset functionality.

Acknowledgment

This research has been funded by the Austrian Research Promotion Agency (FFG) and the Austrian Federal Ministry for Transport, Innovation and Technology (BMVIT), within the "ICT of the Future" project IoT4CPS (Trustworthy IoT for Cyber-Physical Systems) (December 2017 – November 2020).

References

[1] Grieves M. and Vickers J.: "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behaviour in Complex Systems." In: Kahlen J., Flumerfelt S., Alves A. (eds) Transdisciplinary Perspectives on Complex Systems. Springer, Cham (2017)

[2] Oracle: "Digital Twin for IoT Applications: A Comprehensive Approach to Implementing IoT Digital Twin." Oracle White Paper (2017)

[3] Kucera, R., Aanenson, M. and Benson, M.: The Augmented Digital Twin: Combining Physical and Virtual Data to Unlock the Value of IoT. White paper (2017)

[4] Bordawekar, R., Blainey, B., Puri, R.: Analysing Analytics. Morgan & Claypool Publisher (2015)

[5] Geng, X., Arimura, H., and Uno, T. Pattern Mining from Trajectory GPS Data. In: Proceedings of the 2012 IIAI International Conference on Advanced Applied Informatics, IIAIAAI 2012. 60-65 (2012)

[6] Ashbrook, D., Starner, T.: Learning Significant Locations and Predicting User Movement with GPS. In: Proceedings of the IEEE 6th International Symp. on Wearable Comp (2002)

[7] Ashbrook, D., Starner, T.: Using GPS to Learn Significant Locations and Predict Movement Across Multipleusers. In: Proceed. of the Personal and Ubiquitous Comp, 7:275–286 (2003)

[8] Patterson, D., Liao, L., Fox, D., Kautz, H.: Inferring High-Level Behaviour from Low-Level Sensors. In: Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp), Switzerland (2003)

[9] Liao, L., Fox, D., Kautz, H.: Learning and Inferring Transportation Routines. In: Proceedings of the 19th National Conference on Artificial Intelligence (AAAI) (2004)

[10] Han, J., Dong, G., Yin, Y.: Efficient Mining of Partial Periodic Patterns in Time Series Database. In: Proceedings of the International Conference on Data Engineering (ICDE) 1999, Vienna, Austria (1999)

[11] Tsoukatos, I., Gunopulos, D.: Efficient Mining of Spatiotemporal Patterns. In: Proceedings of the International Symposium on Spatial and Temporal Databases (SSTD), USA (2001)

[12] Vlachos, M., Gunopulos, D., Kollios, G.: Discovering Similar Multidimensional Trajectories. In: Proceedings of the International Conference on Data Engineering (ICDE) 2002, pp. 673—684 (2002)

[13] Zheng, V.W., Zheng, Y., Xie, X., Yang, Q.: Collaborative Location and Activity Recommendations with GPS History Data. In: Proceedings of the 19th International Conference on WWW (WWW '10). ACM, USA, 1029-1038 (2010)

[14] Salakhutdinov, R., Mnih, A., Hinton, G.: Restricted Boltzmann Machines for Collaborative Filtering. In: Proceedings of the 24th International Conference on Machine Learning (ICML '07), Z. Ghahramani (Ed.). ACM, New York, NY, USA, 791-798 (2007)

[15] Taylor, G.W., Hinton, G.E.: Factored Conditional Restricted Boltzmann Machines for Modelling Motion Style. In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML '09). ACM, NY, USA, 1025-1032. (2009)

[16] Schölkopf, B., Platt, J., Hofmann, T.: Modelling Human Motion Using Binary Latent Variables, In: Proceedings of the 19th International Conference on Advances in Neural Information Processing Systems 2006, 1, MIT Press, pp.1345-1352 (2006)

[17] Li, X., Zhao, F., Guo, Y.: Conditional Restricted Boltzmann Machines for Multi-Label Learning with Incomplete Labels. In: Proceedings of the 18th International Conference on AI and Statistics, PMLR 38:635-643 (2015)

[18] Lee, J., Kim, H., Kim, N., Lee, J.H.: An Approach for Multi-Label Classification by Directed Acyclic Graph with Label Correlation Maximization. Information Sciences, Vol. 351, pp. 101-114 (2016)

[19] Sutton, R.S. and Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, Cambridge (2012)

[20] Yoo, R.M., Lee, H., Chow, K., Lee, H.S.: Constructing a Non-Linear Model with Neural Networks for Workload Characterization. In: Proceedings of the 2006 IEEE International Symposium on Workload Characterization, San Jose, CA, (2006)

[21] Moreno, I.S., Garraghan, P., Townend, P., Xu, J.: An Approach for Characterizing Workloads in Google Cloud to Derive Realistic Resource Utilization Models. In: Proceedings of the 2013 IEEE 17th International Symposium on Service-Oriented System Engineering (SOSE '13), Washington DC, USA, 49-60. (2013)

[22] Bahga, A., Madisetti, V.K.: Synthetic Workload Generation for Cloud Computing Applications. Journal of Software Engineering and Applications, Vol. 4, pp. 396-410 (2011)

[23] Chen, Y., Ganapathi, A., Griffith, R., Katz, R.H.: Analysis and Lessons from a Publicly Available Google Cluster Trace. In: Proceedings of the EECS Department, University of California, Berkeley, (2010)

[24] Smith, J.W., Sommerville, I.: Workload Classification & Software Energy Measurement for Efficient Scheduling on Private Cloud Platforms. In: Proceedings of the ACM SOCC (2011)

[25] Mishra, A.K., Hellerstein, J., Cirne, W., Das, C.: Towards Characterizing Cloud Backend Workloads: Insights from Google Compute Clusters. In SIGMETRICS Performance Evaluation Review, Vol. 37, pp. 34-41 (2010)

[26] Aggarwal, S., Phadke, S., Bhandarkar, M.: Characterization of Hadoop Jobs Using Unsupervised Learning. In: Proceedings of the Cloud Computing Technology and Science (CloudCom), pp. 748-753 (2010)

[27] Wildstrom, J., Stone, P., Witchel, E., Dahlin, M: Machine Learning for Online Hardware Reconfiguration. In: Proceedings of the 20th International Joint Conference on AI (IJCAI'07), R. Sangal, H. Mehta, and R.K. Bagga (Eds.). Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1113-1118 (2007)

[28] Abbas, A. and Khan, U.S.: A Review on the State-of-the-Art Privacy Preserving Approaches in E-Health Clouds. IEEE Journal of Biomedical and Health Informatics (2014)

[29] Benaloh, J., Chase, M., Horvitz, E., Lauter, K.: Patient Controlled Encryption: Ensuring Privacy of Electronic Medical Records. In: Proceedings of the ACM Workshop on Cloud Computing Security, pp.103-114 (2009)

[30] Chen, T.S., Liu, C.H., Chen, T.L., Chen, C.S., Bau, J.G., Lin, T.C.: Secure Dynamic Access Control Scheme of PHR in Cloud Computing. Journal of Medical Systems, Vol. 36, No. 6, pp. 4005–4020 (2012)

[31] Zhang, R., Liu, L., Xue, R.: Role-Based and Time-Bound Access and Management of EHR Data. Security and Communication Networks (2013)

[32] Bethencourt, J., Sahai, A., Waters, B.: Ciphertext-Policy Attribute-Based Encryption. In: Proceedings of the IEEE Symposium on Security and Privacy SP'07, pp. 321-334 (2007)

[33] Wu, R., Ahn, G.-J., Hu, H.: Secure Sharing of Electronic Health Records in Clouds. In: Proceedings of the 8th IEEE International Conference on Collaborative Computing: Networking, Applications and Work-sharing (CollaborateCom), pp. 711-718 (2012)

[34] Bertino, E., Paci, F., Ferrini,V., Shang, N.: Privacy-Preserving Digital Identity Management for Cloud Computing. IEEE Data Eng. Bull., Vol. 32, No. 1, pp. 21–27, (2009)

[35] Chow, S.S.M., He, Y.-J., Hui, L.C.K., Yiu, S.M.: SPICE – Simple Privacy-Preserving Identity-Management for Cloud Environment. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 526–543 (2012)

[36] Nunez, D., Agudo, I., Lopez, J.: Integrating OpenID with Proxy Reencryption to Enhance Privacy in Cloud-Based Identity Services. In: 4th IEEE International Conference on Cloud Computing Technology and Science Proceedings, CloudCom 2012, pp. 241–248 (2012)+

[37] Nunez, D., Agudo, I.: BlindIdM: A Privacy-Preserving Approach for Identity Management as a Service. International Journal of Information Security, 13(2), pp. 199–215 (2014)

[38] Dey, A., Weis, S.: PseudoID: Enhancing Privacy in Federated Login. Hot Topics in Privacy Enhancing Technologies, pp. 95–107 (2010)

[39] Ardagna, C.A., Camenisch, J., Kohlweiss, M., Leenes, R., Neven, G., Priem, B., Samarati, P., Sommer, D., Verdicchio, M.: Exploiting Cryptography for Privacy-Enhanced Access Control: A Result of the PRIMA Project. Journal of Comp. Sec., 18(1), pp. 123–160 (2010)

[40] Asghar, M.R., Backes, M., Simeonovski, M: PRIMA: Privacy-Preserving Identity and Access Management at Internet-Scale. In: Proceeding of the IEEE International Conference on Communications (ICC) (2016)

[41] Nakrani, P.K.: Smart Car Technologies: A Comprehensive Study of the State of the Art with Analysis and Trends. A MSc Thesis, Arizona State University (2015)

[42] Cyber Security and Resilience of Smart Cars: Good Practices and Recommendations (2016)

[43] Personal Data, Anonymization and Pseudonymization under the GDPR. Salughter and May, https://bit.ly/2tEOHTH

[44] Camenisch, J., Herreweghen, E.V.: Design and Implementation of the Idemix Anonymous Credential System. In: Proceedings of the 9th ACM Conference on Computer and Communications Security CCS '02. New York, NY, USA, pp. 21–30 (2002)

[45] Chaum, D.L.: Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms. Communications of the ACM, Vol. 24, No. 2, pp. 84–90 (1981)

[46] Article 29 Data Protection Working Party: WP29 Guidelines on Automated Individual Decision-Making and Profiling for the Purpose of Regulation 2016/679 (2018)

[47] Sweeney, L.: K-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10(5), 557–570 (2002)

[48] Gkoulalas-Divanis, A., Braghin, S., and Antonatos, S.: FPVI: A Scalable Method for Discovering Privacy Vulnerabilities in Microdata. In: Proceeding of the *2016 IEEE International Smart Cities Conference (ISC2),* 1-8 (2016).