Extending Model-Based Privacy Analysis for the Industrial Data Space by Exploiting Privacy Level Agreements

Amir Shayan Ahmadian
University of Koblenz Landau, Germany
ahmadian@uni-koblenz.de

Jan Jürjens
University of Koblenz Landau, Germany
Fraunhofer ISST, Germany
http://jan.jurjens.de

Daniel Strüber
University of Koblenz Landau, Germany
strueber@uni-koblenz.de

ABSTRACT

Considering the dramatic impact of the current technology changes on user privacy, it is important to contemplate privacy early on in software development. Ensuring privacy is particularly challenging in industrial ecosystems, in which an enterprise may depend on or cooperate with other enterprises to provide an IT service to a service customer. An example for such ecosystems is the Industrial Data Space (IDS). The IDS provides a basis for creating and using smart IT services, while ensuring digital sovereignty of service customers. In this paper, motivated by Article 25 of Regulation (EU) 2016/679 (GDPR), we apply a model-based privacy analysis approach to the IDS to enable the verification of conformance to customer's privacy preferences. To this end we extend an existing model-based privacy analysis to support customer's privacy preferences in compliance with the Article 5 of the GDPR. We also provide a privacy check to support the privacy of data exchanges between the enterprises. The approach is supported by the CARiSMA tool.

CCS CONCEPTS

• **Security and privacy** → *Software security engineering*; • **Software and its engineering** → *Software design engineering*;

KEYWORDS

Privacy by Design, Model-based Privacy Analysis, Industrial Data Space, Personal Data, GDPR

ACM Reference Format:
Amir Shayan Ahmadian, Jan Jürjens, and Daniel Strüber. 2018. Extending Model-Based Privacy Analysis for the Industrial Data Space by Exploiting Privacy Level Agreements. In *SAC 2018: SAC 2018: Symposium on Applied Computing , April 9–13, 2018, Pau, France.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3167132.3167256

1 INTRODUCTION

Privacy has recently become a major factor in any kind of software development [25]. Nowadays, most of the enterprises that provide IT services require the personal information of their service customers, who provide data, to perform their business processes. As a result, an enormous amount of personal data is collected, stored, and shared all over the world [12]. Failure to protect such data by enterprises affects the data providers (service customers) negatively, and may harm the reputation of service providers (enterprises) and cause emotional or financial damages.

Article 25 of the EU General Data Protection Regulation (GDPR) prescribes privacy by design (PbD) [33]. This requires that the IT systems must be focused or technically adapted for ensuring that, by default, the principles relating to the processing of personal data are supported. According to Cavoukian [10], who first introduced Privacy by design (PbD), PbD implies that the design of a system must be analyzed regarding legal requirements and customers' privacy preferences, and where necessary, it must be improved—for instance by integrating privacy enhancing technologies—to support privacy. Despite the availability of a range of privacy enhancing technologies (PETs) [5, 14, 15], which provide strong privacy guarantees in different domains, according to Spiekermann [30, 31], PbD is a powerful term, and include more than the process of uptaking a few PETs. Prior to the integration of PETs into a system, the system design must be analyzed in early phases of the system development.

System-level privacy analysis is particularly challenging in today's digital society, where industrial ecosystems play a key role. An enterprise may depend on or cooperate with other enterprises to provide an IT service to a service customer. An example for such ecosystems is the Industrial Data Space (IDS) [6]. The IDS aims at establishing a network for trusted data exchange between different enterprises, which provides or consumes (processes) data.

A strategic requirement of the IDS is to provide secure data supply chains, to ensure a high level of confidence when exchanging and processing data. The current reference architecture of the IDS (provided in [6]) does not consider privacy explicitly. In particular, it does not specify mechanisms to ensure that the principles on processing of personal data introduced in Article 5 of the GDPR are respected.

In [3] a model-based privacy analysis approach is introduced. This approach generally enables one to verify if the design of a system that processes personal data supports the privacy preferences of the service customers. The privacy preferences are based on the key elements of privacy introduced in Barker et al.'s seminal taxonomy [7]: *purpose, visibility, granularity*, and *retention.* In nutshell, in their approach the authors verify if: (I) a piece of personal data is processed for a set of specific authorized reasons (purposes), (II) the access to personal data is restricted to authorized persons, (III) the

SAC 2018, April 9–13, 2018, Pau, France

ACM ISBN 978-1-4503-5191-1/18/04…$15.00
https://doi.org/10.1145/3167132.3167256

granularity level of personal data is respected when personal data are sent to third parties, (IV) restrictions mechanisms are available to ensure that personal data are eventually deleted or restricted. We leverage this approach to perform a privacy analysis on the IDS to verify if the privacy preferences of the data providers are supported.

However, the reference architecture of the IDS [6] differs from the architecture analyzed in [3]. In the IDS, the exchange of data is enabled through *connectors*, that is, dedicated communication servers for sending and receiving data. Moreover, Article 5 of the GDPR stipulates 6 principles on processing personal data, which correspond to the four key privacy elements mentioned above. In these 6 principles—in contrast to the approach presented in [3]—among the four privacy elements, *purpose* is considered as the fundamental element, and the other three elements are defined in relation with the purpose. For instance, personal data must be kept no longer than is necessary for the legitimate purposes, or only authorized persons may have access to data for authorized purposes.

In this paper we make the following main contributions: (I) We highlight the importance of addressing privacy of personal data in the reference architecture of the IDS. (II) We explain how agreements that specify the principles on processing of personal data may be established between data providers and data consumers to support the privacy analysis in the IDS. (III) We extend the model-based privacy analysis introduced in [3] and apply the adapted analysis to the IDS. (IV) We validate our extended model-based privacy analysis concerning the privacy targets introduced in [24].

The paper is organized as follows. In Section 2, the necessary background is provided. In Section 3, we describe the privacy challenges regarding the IDS. In Section 4, we extend the model-based privacy analysis and demonstrate its application to the IDS. In Section 5, we investigate the related work. In Section 7, we conclude.

2 BACKGROUND

Nowadays, system models are widely used for formal or informal (for learning or communication) purposes, in industry [32]. System models assist system developers to handle the complexity of systems by means of abstraction and focus on main concerns such as privacy [13].

An overview of the model-based privacy analysis approach presented in [1, 3] is provided in Figure 1. The key idea is to exploit privacy level agreements (PLAs). PLAs are appendixes to service level agreements, and offer a structured way to communicate the privacy protection level of personal data provided by a service provider to a service customer [11]. In nutshell, given a PLA (which contains privacy preferences of a service customer) and a system model, four privacy checks are provided to analyze the system model according to the privacy preferences to generate a set of analysis results. Systems are modeled using UML, and as a basis to implement the privacy checks, two UML extensions are introduced: the *privacy* and *rabac* profiles. The former one is used to express privacy-specific information in a system model. The later one uses the *role- and attribute-based access control* model [19] to enable the software designers to identify who is authorized to process personal data concerning the structure and the behavior of a system.

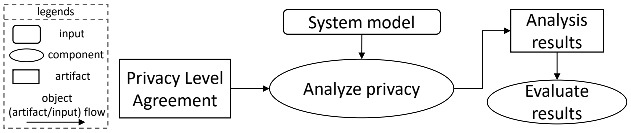

Figure 1: Model-based privacy analysis by exploiting PLAs.

In this paper we use different stereotypes of these two UML extensions to annotate the UML diagrams in the context of the IDS. The stereotype ≪sensitiveData≫ is used to specify that a piece of data relates to an identified or identifiable natural person and according to the GDPR is a piece of personal data. The stereotype ≪recipient≫ is used to specify that a *NamedElement* such as a *dataStore* node in an activity diagram belongs to an enterprise (a data controller or data processor) to which a piece of personal data is transferred. The stereotype ≪granularity≫ is used to specify the level of the precision of a piece of personal data. The stereotype ≪objective≫ is used to specify the processing purposes of an operation of a class. The stereotypes, ≪abac≫ and ≪abacRequire≫ are used to specify different subjects (persons), their roles and rights in a system.

3 ADDRESSING PRIVACY IN THE IDS

The terms used in this paper are based on the terms and definitions of the GDPR. According to the GDPR, a *data controller* determines the purposes and the means for the processing of personal data (privacy preferences). Concerning the IDS, a *data provider* is a data controller, who provides data (including personal data) and specifies the privacy preferences of these data. According to the GDPR, a *data processor* processes personal data on behalf of the controller. Concerning the IDS, a *data consumer* either refers to a data processor, who directly processes the provided data, or a data controller is denoted, who transfers to other data processors the data and their privacy preferences specified by the data provider.

The IDS initiative was launched in Germany by representatives from business, politics, and research. The aim is to provide a virtual data space for secure data exchanges. Currently the IDS includes 68 organizations. It establishes secure data supply chains from data source to data use, while ensuring data sovereignty for data providers [6, 26]. It aims to provide a technology which is simple, reliable, and cheap for every citizen zu use which that preserves the digital sovereignty of the citizen. In particular, the goal is to provide a platform for collaborative smart data analytics which supports a true Digital Sovereignty of the private data of the citizens in order to put them in a sustainable position to control who receives their personal data and what they can do with it.

The main activities of the IDS are: (I) **Providing data** is enabled through the *Broker* service. The *Broker* service indexes the metadata that is provided by a data provider (data controller). The metadata describe the source of data, and contain a set of policies on using the data. (II) **Exchanging data** is initiated by a data consumer requesting data from a broker. The request and the exchange of data is enabled by the IDS *connectors* that are deployed on each enterprise. (III) **Data Processing** is performed by the data applications and enterprises' services.

A strategic requirement of the IDS is to ensure a high level of confidence during data exchange. To this end, the IDS reference architecture requires the use of a security profile in order to implement appropriate mechanisms to ensure secure data communication between connectors, provide proper access control mechanisms to support identity and access management, and make use of cryptographic methods to establish trust across the entire business ecosystem and protect the IDS participants from fraud.

Article 5 of Regulation (EU) 2016/679 stipulates six principles for the processing of personal data: personal data must be (a) processed lawfully, (b) collected for specified and legitimate *purposes*, (c) adequate and limited to what is necessary regarding the purposes (*granularity*), (d) accurate and kept up to date, (e) kept no longer than is necessary for the authorized purposes (*retention*), and (f) protected against unauthorized processing (*visibility*). The current security profile of the IDS does not require the use of mechanisms to ensure that the personal data processing in the IDS respects these principles. For instance, there is currently no mechanism prescribed to be used to ensure that personal data is only processed for a certain set of processing purposes, or the stored personal data in a database of a data consumer are eventually deleted or restricted, or during personal data exchange the granularity levels are respected.

The usage scenarios of the IDS span a large variety of domains, including automotive engineering, facility management, healthcare, and smart cities [18]. To illustrate the need for data protection in the IDS, consider the following concrete usage scenarios. (I) Sensors embedded in *car seats*: Such sensors are designed to improve the ergonomics of a smart car. The data produced by these sensors are transmitted to central monitoring systems (through the IDS connectors) and stored in different databases. Such data may reveal physiological aspects of a car driver (for instance, by transmitting her/his weight average). (II) Sensors embedded into infrastructural objects (for instance *trash cans*) to support smart services in smart cities: These trash cans may be managed by different operatives. The sensors embedded in these trash cans may log information about the *who* and *when* of trash can uses, and through the IDS connectors transmit such logged information. Such information may reveal the time schedule of the operatives. For instance when the operatives work or have breaks.

According to these two scenarios, the data that are exchanged between connectors may include some information about individuals. This makes it necessary to analyze the system's design of the connectors (as the central functional entity of the IDS) to verify if the principles on the processing of personal data are supported. In the following section, to fully support the privacy principles prescribed in Article 5 of the GDPR, we first extend the model-based privacy analysis introduced in 2, and then apply the extended privacy analysis to the IDS in order to ensure privacy protection in the early phases of system development.

4 MODEL-BASED PRIVACY ANALYSIS FOR THE IDS

In this section, we first define privacy preferences in the IDS in compliance with the GDPR, and describe how such privacy preferences are specified for a piece of personal data. Afterwards, we extend the model-based privacy analysis introduced in [3], regarding the reference architecture of the IDS, and the definition of the privacy preferences.

4.1 Privacy Preferences

The security profile of the IDS manifests some high-level attributes such as hardware security, access controls, and authentication level [6]. To support privacy principles, the security profile of IDS must particularly specifies the personal data that are processed in the IDS. Moreover, a set of preferences on the processing of personal data in the IDS must be defined. These preferences are based on the four fundamental privacy elements introduced in [7].

- **Purpose** is the basic element of data privacy. It indicates the authorized reasons to process data.
- **Visibility** indicates who is allowed to process the data provided for an authorized purpose.
- **Granularity** refers to characteristics of data that could be used to facilitate proper processing of the data, when different valid accesses for various purposes exist. In other words, data granularity specifies how much precision is provided in response a query.
- **Retention** refers to the need to restrict processing or removing the data after they have been processed for the intended purposes.

In the definition of the privacy preferences in [3] the four key privacy elements, namely *purpose*, *visibility*, *granularity*, and *retention* are considered separately. However, as we mentioned before, according to Article 5 of the GDPR, *purpose* is the fundamental element in privacy and other three privacy elements are defined in regard to *purpose* (which is not the case in [3]). Therefore, in this paper we define the privacy preferences as:

Definition 4.1. Let P be a partially ordered set of all defined purposes, V be a partially ordered set of all subjects who can process a piece of personal data, G be a set of all granularity levels and R be a set of retention conditions. The privacy preferences (PRP) of a piece of personal data pd is defined:

$$PRP_{pd} = P \nrightarrow \mathcal{P}(V) \times G \times R$$

According to this definition, a partial function maps a *purpose* $p \in P$ to a subset of all possible subjects who are allowed to process data (the power set of V), a granularity level $g \in G$, and a retention condition $r \in R$. Instead of a function, a partial function is used to state that not for all possible purposes a mapping is required. In other words, not necessarily all purposes are defined as authorized purposes, and only a subset of all possible purposes may be contained in the privacy preferences.

For instance consider the four simple sets P={marketing, assess, invoice}, V={finAd, markM, techM, techAd}, G={none, existential, partial, specific}, R={1M, 1Y}. Then for a piece of personal data such as a *credit card number (ccn)* the privacy preferences may be defined as:

$$PRP_{ccn} = \{(invoice \mapsto ((finM, finAd), existential, 1Y))\}$$

PRP_{ccn} specifies that a credit card number may be processed for the purpose of *invoice*. For this purpose a *financial manager* (*finM*) and a *financial administrator* (*finAd*) are allowed to process it for a

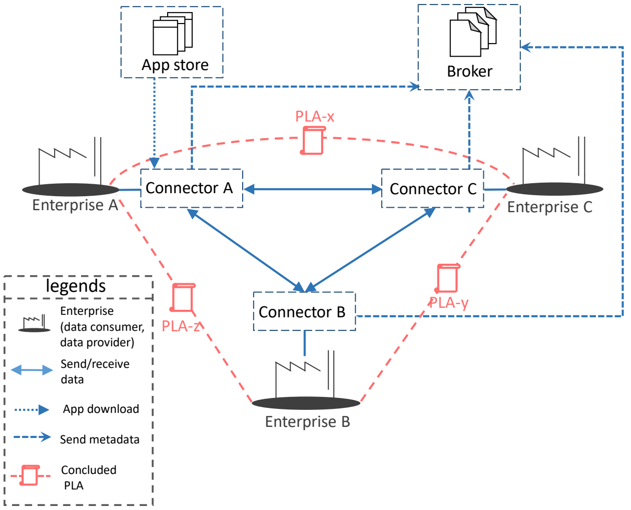

Figure 2: An illustration of the IDS system layer, including privacy level agreements

period of *1 year*. The *ccn* may be transferred to other enterprises for the purpose of *invoice* by only stating that such data exist and not provide the complete (specific) credit card number.

Similar to the definition of privacy preferences in [3], in Definition 4.1, the purpose (P) and data subject sets (V) are partial ordered sets. This enables one to define lattice structures, where each node corresponds to a member from these two sets, and each edge demonstrates a hierarchical relation between two members which subsume each other. For instance in a lattice that organizes the set V={finAd, markM, techM, techAd}, *technical manager (techM)* is the descendent (child) of *technical administrator (techAd)*.

In the IDS, we specify the privacy preferences for each piece of personal data in privacy level agreements (PLAs). Between each two enterprises that exchange data in the IDS a PLA is concluded. Additionally, in a PLA some specific information on each enterprise such as enterprise identity and representative(s) are included. Figure 2 illustrates an excerpt from the IDS system layer, including three enterprises and the concluded PLAs between them. Connectors are communication servers for sending and receiving data. In each enterprise, data (including personal data) are processed by applications that are deployed on each connector. These applications are either downloaded from the *App Store* of the IDS or are self-developed apps.

The personal data processing in each enterprise must support the privacy preferences included in PLAs. In the following section we describe how a privacy analysis is performed in the IDS to verify if the privacy preferences are supported.

4.2 Extending model-based privacy analysis

According to the reference model of the IDS, to ensure privacy of personal data, data processing should be performed as close as possible to the data source, rather than be delegated to other enterprises. If the data (including personal data) are intended to be transferred to external enterprises, the data processing on an external enterprise must respect the privacy preferences specified in the PLA concluded between the two enterprises (the data provider and the data consumer). To verify if the privacy preferences are supported in this case, the system design of the enterprise, to which personal data are sent, must be analyzed. We use a model-based approach to perform such a privacy analysis.

Connectors are the central functional entity of the IDS for exchanging and processing data. Independent of the apps being deployed on the connectors, a system model including several UML diagrams—in particular, *class, activity, component* and *deployment* diagrams—describes the structure and a behavior of a connector. Such a system model belongs to the *configuration model* of a connector. According to the IDS, a configuration model describes the configuration of a connector in a technology-independent manner. Since the existing system models of the connectors are specified using UML, they are amenable to the model-based privacy analysis introduced in [3] (see Section 2).

4.2.1 Privacy check.

In [3], four privacy checks are introduced to analyze a system model concerning the four key privacy elements, namely *purpose*, *visibility*, *granularity*, and *retention*, to verify if the privacy preferences are supported. These four privacy checks are performed independently. In Section 4.1, we provided a new definition for the privacy preferences. Since in this definition *purpose* is the fundamental privacy element and other three elements are defined in relation to it, our privacy analysis differs from the privacy analysis introduced in [3]. In a system model, we first verify if a piece of personal data is processed for the authorized purposes. Afterwards, we verify the system model considering the *visibility*, *granularity*, and *retention* elements, in relation to the authorized purposes. In fact, we only perform one single privacy check instead of four separate checks.

To perform a privacy analysis on a system model, the system model must be annotated with privacy elements. The privacy profile described in Section 2 is used to annotate a system model. Given a system model of a connector deployed on a enterprise, we first identify the processing purposes of the operations that process personal data. The processing purposes of the operations will be compared with the authorized purposes specified in the privacy preferences included in a PLA. If a piece of personal data is processed for unauthorized purposes, the privacy analysis fails and a report notifying this failure will be generated. If there is no conflict between the processing purposes of the system operations and the authorized purposes, the privacy analysis considering each authorized purpose for which a piece of personal data is processed in a system, verifies the following: (I) if the subjects (persons) who process the piece of personal data, are actually authorized to do this. (II) If the piece of personal data is transferred to another enterprise for the authorized purpose, the granularity level is respected by the data transmission. (III) if the piece of personal data is stored in an enterprise for the authorized purpose, an operation(s) exists that deletes or restricts the stored piece of personal data. If any of these three cases fails, a report notifying the failure will be generated.

In a system model, the activity diagrams model the processing of data objects. The class diagrams model the structure of the system. A piece of data is an object (in an activity diagram), or a class (in a class diagram).

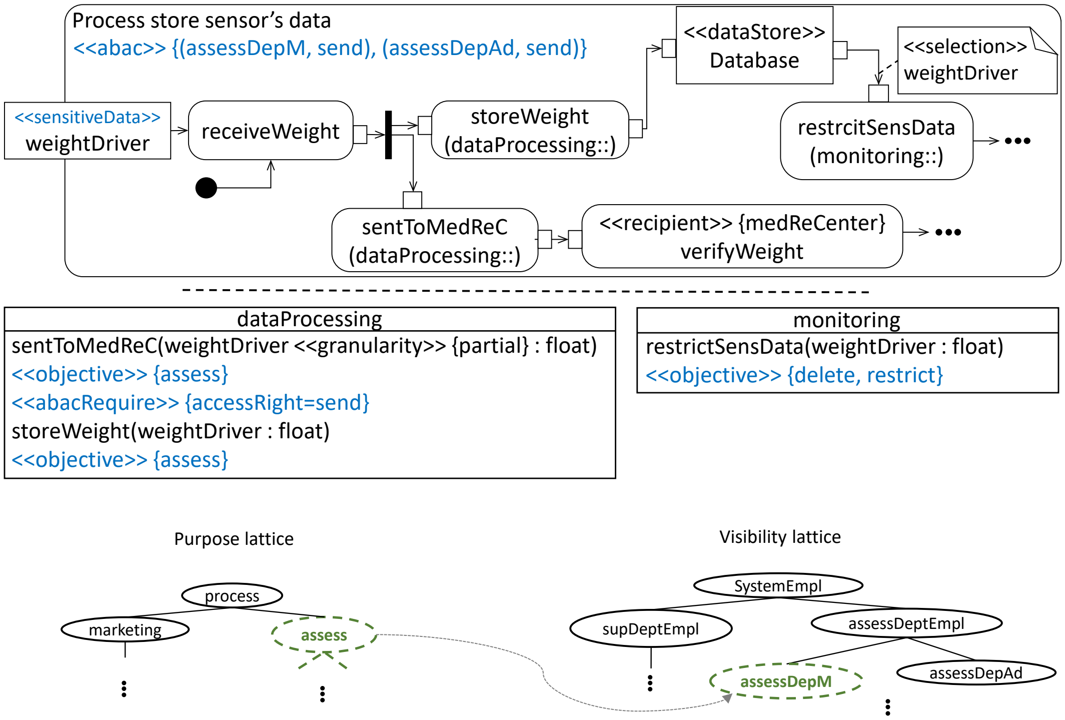

Figure 3: Design model excerpt annotated with privacy profile, demonstrating the privacy analysis. The privacy preferences are presented as dashed-lines in the respective lattices.

Concerning the usage scenario on the sensors embedded in car seats, in Figure 3, excerpts of an activity and a class diagram are provided. The activity diagram models the process of receiving and storing the weight of a car driver by a monitoring system. Moreover in this activity, the weight is transferred to a *medical research center* for further assessments and comparing with the existing medical data stored in the research center. This information are sent from the sensors that are embedded in car seats. As we previously mentioned, such information reveals physiological aspects of a car driver. Hence, as illustrated in the activity, the object node is annotated with ≪sensitiveData≫. The *verifyWeight* action in the activity diagram is annotated with the stereotype ≪recipient≫ {medReCenter} specifying that this action corresponds to an operation which belongs to the system model of the *medical research center*.

The operations in the classes are annotated with the stereotype ≪objective≫ and the relevant tags to express the processing purposes of each operation. The parameter of the operation *sendToMedReC* is annotated with ≪granularity≫ {partial}, specifying that the required precision level of the piece of personal data to be transferred to a recipient is partial. Moreover, the stereotype ≪abacRequire≫ specifies the access rights of an operation. Using the access rights and concerning the stereotype ≪abac≫, the subjects who process a piece of personal data are identified. For instance, according to Figure 3, both the *assessment department manager* (*assessDepM*), and *administrator* (*assessDepAd*) process (send) the weight of a car driver to a recipient.

Moreover, in Figure 3 excerpts of 2 lattices are demonstrated, namely a *purpose-lattice*, and a *visibility-lattice*. The former specifies a set of all possible purposes. The latter specifies a set of all subjects, who are authorized to process a piece of data for a specific purpose. Consider that the following privacy preferences for the weight of a car driver (*wcd*) is specified in the PLA concluded between the enterprise that gather the information from the seat sensors and the enterprise that monitor and assess such information:

$$PRP_{wcd} = \{(assess \mapsto (assessDepM, partial, 1M))\}$$

In the lattices, the privacy preferences are specified with the dashed-lines, and specify that the weight of a car driver may be processed by the assessment-department manager (assessDepM) for the purpose of assessment for the period of one month. If the weight is transferred to other enterprises for further processing, the precision level must be partial (for example a range for the weight must be provided).

According to our privacy check, to analyze the model provided in Figure 3, we first compare the authorized purpose specified in the privacy preferences with the processing purposes (objectives). Two actions in the activity diagram process the weight of a car driver, namely *sendToMedReC* and *storeWeight*. Concerning the respective operations provided in the class *dataProcessing*, processing purpose is *assess*, which complies with the authorized purpose in the privacy preferences. We now, compare the visibilities. According to the privacy preferences, only an assessment-department manager (*assessDepM*) is authorized to process the weight of a driver for the purpose *assess*. However, according to Figure 3, in addition to *assessDepM*, an administrator can also process the weight. This is a violation, thus, the privacy check fails. Regarding the granularity level, the model complies with the privacy preferences. Furthermore, in the system model there exists an operation with the objective *delete* and *restrict*, that will eventually—when the processing for the purpose of *assess* is finished—deletes or restricts the weight of a car driver, from the data base (*dataStore* node) of the monitoring system.

4.2.2 Broker-check. According to the system layer of the IDS in Figure 2, an enterprise through its connectors may exchange data with an IDS broker. Such a data exchange is enabled through metadata, which describe the source of data, and provide a set of policies on using the data. A data exchange with an IDS broker must not contain personal data. Metadata only aim to initiate data exchanges between IDS connectors.

We introduce a new check (*broker-check*) to ensure that a data exchange between an enterprise and an IDS broker does not include personal data. Given an activity diagram which models the data exchange with an IDS broker, the metadata that are sent to a *namedElement*—such as a *dataStore* node—annotated with ≪recipient≫ {IDS-broker} must not be annotated with ≪sensitiveData≫.

Figure 4 shows an excerpt from an activity diagram specifying the process of sending metadata from a connector to an IDS broker. The *dataStore* node is annotated with ≪recipient≫ {*IDSbroker*} specifying that this node is a database in an IDS broker *IDSbroker*. The object which is sent to this *dataStore* is annotated with ≪sensitiveData≫. This is a violation, and the broker-check fails.

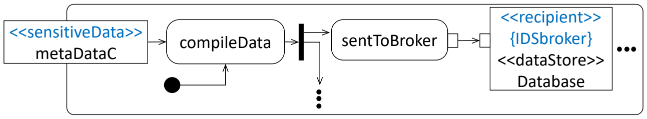

Figure 4: Design model excerpt annotated with privacy profile, demonstrating broker-check

5 DISCUSSION

Since the system models of the IDS must be treated in confidence, we do not provide the actual system models of the IDS in this paper. The design model excerpt presented in Figure 3 is based on the existing system models (UML) and example scenarios of the IDS. Generally, by applying the model-based privacy analysis to the IDS (the car seat's sensors scenario), we noticed that such an analysis can successfully support Privacy by Design in the IDS. The identification of violations—which specify that a system model is not fully in compliance with a set of privacy preferences—assists practitioners to support privacy requirements in the early phases of system development, and facilitates the integration of privacy enhancement technologies (PETs) into the system design.

Particularly in this paper, (I) We described the importance of addressing the privacy of personal data in the reference architecture of IDS. For this, we described two example scenarios from the IDS, in which failures to ensure privacy protection may affect the data providers and the data consumers. (II) According to Article 5 of

Table 1: The privacy targets supported by the IDS and the privacy check described in Section 4.2

Privacy target	
P1. Data quality	
P1.1 Ensuring fair and lawful processing by transparency	PC
P1.2 Ensuring processing only for legitimate purposes	PC
P1.3 Providing purpose specification	PC
P1.4 Ensuring limited processing for specified purposes	PC
P1.5 Ensuring data avoidance	IDS
P1.6 Ensuring data minimization	PC
P1.7 Ensuring data quality, accuracy and integrity	IDS
P1.8 Ensuring limited storage	IDS
P2. Ensuring legitimacy	
P2.1 Ensuring legitimacy of personal data processing	PC
P2.2 Ensuring legitimacy of sensitive personal data processing	PC
P3. Providing adequate information	
P3.1 Adequate information in case of direct collection of data	PC
P3.2 Adequate information where data is not obtained directly	IDS
P4. Access right of data subject	
P4.1 Facilitating the provision of information about processed data and purpose	PC
P4.2 Facilitating the rectification, erasure or blocking of data	PC
P4.3 Facilitating the portability of data	IDS
P4.4 Facilitating the notification to third parties about rectification, erasure and blocking of data	IDS
P5. Data subject's right to object	
P5.1 Facilitating the objection to the processing of data	PC
P5.2 Facilitating the objection to direct marketing activities	PC
P5.3 Facilitating the objection to data-disclosure to others	PC
P5.4 Facilitating the objection to decisions on automated processing	IDS
P5.5 Facilitating the data subjects right to dispute the correctness of machine conclusions	IDS
P6. Security of data	
P6.1 Ensuring the confidentiality, integrity, availability, and resilience	IDS
P6.2 Ensuring the detection of personal data breaches and their communication to data subjects	IDS, PC
P7. Accountability	IDS
P7.1 Ensuring the accountability	

PC:	Privacy check described in this paper
IDS:	Supported by the IDS

Regulation (EU) 2016/679, we provided a new definition for privacy preferences. The privacy preferences were originally defined in [3], however, we noticed that in their definition *purpose* is not defined as the fundamental privacy element. We further explained how such preferences are specified for a piece of personal data in PLAs that are concluded between enterprises to support model-based privacy analysis in the IDS. (III) We extended the model-based privacy analysis introduced in [3] to support our new definition of the privacy preferences and the reference architecture of the IDS. Except the *broker check*, our extended privacy analysis is not limited to the IDS and may be used to analyze other IT system models. (IV) We applied the extended privacy analysis to a system model derived from the IDS.

The privacy analysis is supported by an open source tool called CARiSMA[1] [2], which is a platform-independent tool. It is originally designed to make a security analysis based on the UMLsec profile available to developers [21, 28]. Due to its flexible plugin architecture, we extended CARiSMA to support our privacy check.

In [24], the authors provide a systematic support for representing privacy requirements in the form of privacy targets. The privacy targets are derived from legal privacy and data protection principles. The privacy targets provided in [24], support the privacy principles prescribed in the GDPR. To validate the model-based privacy analysis introduced in this paper, we verify how these privacy targets are supported through the application of our privacy analysis approach to the IDS (Table 1).

A number of the privacy targets are supported by the IDS. For instance, *accountability*, *security of data*, and *data accuracy and integrity* are supported by the security profile of the IDS. Moreover, the IDS provides appropriate mechanisms to ensure *limited storage*, *data portability*, and *notifications to the third party* [6].

The IDS does not prescribe mechanisms to support the privacy targets that are related to the privacy elements, namely *purpose*, *visibility*, *granularity*, *retention*. Our privacy analysis provides a mechanism to analyze a system model of the IDS to verify if the privacy elements as well as the relevant privacy targets are supported by a system. For instance, the specification of authorized purposes in a PLA and their comparison with the processing purposes of a system, support the privacy targets *P1.2 - P1.4*, *P1.6*, *P3.1*, and *P5.2*.

Lessons learned: We noticed that a model-based analysis does not identify all privacy issues of a system. For instance, concerning retention, at design time we are not able to verify if a mechanism is triggered exactly by expiring the retention time to delete or restrict a piece of personal data. However, model-based techniques assist software developers to be aware of privacy issues from the early phases of the software development and support them in their implementation.

Moreover, modeling a system does not mean that the code of the system completely conforms to the system model. Therefore, analyzing a system model does not guarantee that the privacy requirements are completely supported in the source code. Using reverse engineering, the code of a system may be transformed to a system model on which privacy analysis may be performed.

6 RELATED WORK

Privacy by design as a concept has existed for years, however, only just becoming part of a legal requirement with the GDPR. A number of methodologies and approaches are proposed to support and manage privacy issues in system design. However, the majority of these techniques either do not provide an approach to support privacy from early design phases, or do not address privacy specifically.

UMLsec [21] provides an approach to develop and analyze security critical software, in which security requirements such as integrity, availability, and confidentiality are specified in system models [17, 29]. UMLsec implementation has been applied to various industrial applications [20, 22]. However, UMLsec analysis does not consider privacy.

In [8], a UML profile for privacy-aware applications is provided. This profile enables one to describe a privacy policy that is applied by an application and keep track of which elements are in charge of enforcing it. They apply their approach to the privacy policy of Google services. This profile does not enable one to analyze a system design.

In [23], a method (PriS) for incorporating privacy requirements into the system design process is proposed. The PriS method enables the analysis of the effect of privacy requirements on organizational processes. This method is applied to two case studies, an e-voting system and a career office system of a university. The PriS method can be used to effectively link organizational privacy needs to system implementation and can guide designers to make proper decisions regarding the most appropriate technological solution. However, this method does not enable one to perform a privacy analysis to verify if privacy requirements are supported.

In [27], a variety of guidelines and techniques to assist practitioners and software engineers to support privacy when the systems are designed. This approach allows enterprises to determine whether system configurations or processes do actually conform to their assertions about privacy-respecting safeguards. However, this approach is rather generic and the authors only focus on top-level privacy goals.

In [9] and it's recent variation [16], the authors developed a framework for privacy-aware design in the field of ubiquitous computing. In this framework, a set of questions is provided that enables the designers to evaluate a system. Although these frameworks are fast to implement and inexpensive, they only identify a set of static privacy problems in systems, and no privacy analysis on the system's design is performed.

7 CONCLUSION

The work presented in this paper is motivated by Article 25 of the GDPR. We explained the importance of supporting privacy in the Industrial Data Space (IDS). We provided a definition for privacy preferences in compliance with Article 5 of the GDPR. Concerning this definition, we extended the existing model-based privacy analysis. Furthermore, to support the privacy of data exchange in the IDS we introduced a new check in addition to the extended privacy analysis. We discussed the results of the application of the model-based privacy analysis to the IDS concerning the privacy targets that are derived from the GDPR. Analyzing the design of

[1] http://carisma.umlsec.de

a system using a model-based approach in the early phases of the system development facilitates Privacy by Design.

In our ongoing work, we investigate how we may evaluate analysis results to identify privacy threats and support a privacy risk assessment using a model-based analysis [4]. In our future work, we will investigate how a privacy analysis on abstract system models may be performed, and how the conflicts between a system design and a set of privacy preferences may be handled.

ACKNOWLEDGMENTS

This research was partially supported by: (I) Design For Future – Managed Software Evolution (DFG's SPP 1593, project JU 2734/2-2), (II) Engineering Responsible Information Systems (University of Koblenz Landau), (III) Industrial Data Space, German Ministry of Research no. 01IS15054.

REFERENCES

[1] Amir Shayan Ahmadian and Jan Jürjens. 2016. Supporting Model-Based Privacy Analysis by Exploiting Privacy Level Agreements. In *2016 IEEE International Conference on Cloud Computing Technology and Science, CloudCom 2016, Luxembourg, December 12-15, 2016*. 360–365.

[2] Amir Shayan Ahmadian, Sven Peldszus, Qusai Ramadan, and Jan Jürjens. 2017. Model-based privacy and security analysis with CARiSMA. In *Proceedings of the 2017 11th Joint Meeting on Foundations of Software Engineering, ESEC/FSE 2017, Paderborn, Germany, September 4-8, 2017*. 989–993. https://doi.org/10.1145/3106237.3122823

[3] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. 2017. Model-Based Privacy Analysis in Industrial Ecosystems. In *Modelling Foundations and Applications - 13th European Conference, ECMFA 2017, Held as Part of STAF 2017, Marburg, Germany, July 19-20, 2017, Proceedings*. 215–231. https://doi.org/10.1007/978-3-319-61482-3_13

[4] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. 2018. Supporting Privacy Impact Assessment by Model-Based Privacy Analysis. In *Proceedings of ACM SAC Conference (SAC18)*. ACM, New York, NY, USA. accepted.

[5] Thibaud Antignac and Daniel Le Métayer. 2014. *Privacy by Design: From Technologies to Architectures*. Springer International Publishing.

[6] S. Auer, J. Jürjens, B. Otto, G. Brost, C. Lange, C. Quix, J. Cirullies, S. Lohmann, J. Schon, A. Eitel, C. Mader, D. Schulz, T. Ernst, N. Menz, J. SchÃijtte, C. Haas, R. Nagel, M. Spiekermann, M. Huber, H. Pettenpohl, S. Wenzel, C. Jung, and J. Pullmann. 2017. *Reference architecture model for the industrial data space*. White Paper. Fraunhofer. www.industrialdataspace.org

[7] Ken Barker, Mina Askari, Mishtu Banerjee, Kambiz Ghazinour, Brenan Mackas, Maryam Majedi, Sampson Pun, and Adepele Williams. 2009. *A Data Privacy Taxonomy*. Springer Berlin Heidelberg.

[8] T. Basso, L. Montecchi, R. Moraes, M. Jino, and A. Bondavalli. 2015. Towards a UML Profile for Privacy-Aware Applications. In *2015 IEEE International Conference on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*.

[9] Victoria Bellotti and Abigail Sellen. 1993. Design for Privacy in Ubiquitous Computing Environments. In *Proceedings of the Third Conference on European Conference on Computer-Supported Cooperative Work*.

[10] Ann Cavoukian and Michelle Chibba. 2009. Advancing privacy and security in computing, networking and systems innovations through privacy by design. In *Proceedings of the 2009 conference of the Centre for Advanced Studies on Collaborative Research, November 2-5, 2009, Toronto, Ontario, Canada*.

[11] Cloud Security Alliance. 2013. Privacy Level Agreement [V2]: A Compliance Tool for Providing Cloud Services in the European Union. (2013).

[12] Pietro Colombo and Elena Ferrari. 2012. Towards a modeling and analysis framework for privacy-aware systems. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Confernece on Social Computing (SocialCom)*. IEEE, 81–90.

[13] Robert France and Bernhard Rumpe. 2007. Model-driven Development of Complex Software: A Research Roadmap. In *2007 Future of Software Engineering (FOSE '07)*.

[14] Munawar Hafiz. 2013. A pattern language for developing privacy enhancing technologies. *Softw., Pract. Exper.* 7 (2013), 769–787.

[15] Jaap-Henk Hoepman. 2014. *Privacy Design Strategies*. Springer Berlin Heidelberg.

[16] Jason I. Hong, Jennifer D. Ng, Scott Lederer, and James A. Landay. 2004. Privacy Risk Models for Designing Privacy-sensitive Ubiquitous Computing Systems. In *Proceedings of the 5th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '04)*.

[17] S. H. Houmb, G. Georg, J. Jürjens, and R. B. France. 2006. An Integrated Security Verification and Security Solution Design Trade-off Analysis Approach. In *Integrating Security and Software Engineering: Advances and Future Vision*, H. Mouratidis (Ed.). Idea Group, 190–219.

[18] Industrial Data Space Association. 2014. Industrial Data Space Use Case Broschuere. (2014).

[19] Xin Jin, Ravi Sandhu, and Ram Krishnan. 2012. *RABAC: Role-Centric Attribute-Based Access Control*. Springer Berlin Heidelberg.

[20] J. Jürjens. 2001. Modelling audit security for smart-card payment schemes with UMLsec. In *Trusted Information: The New Decade Challenge*. Proceedings of the *16th International Conference on Information Security (SEC 2001)*.

[21] Jan Jürjens. 2005. *Secure systems development with UML*. Springer.

[22] J. Jürjens and G. Wimmel. 2001. Formally Testing Fail-safety of Electronic Purse Protocols. In *16th International Conference on Automated Software Engineering (ASE 2001)*. IEEE, 408–411.

[23] Christos Kalloniatis, Evangelia Kavakli, and Stefanos Gritzalis. 2008. Addressing privacy requirements in system design: the PriS method. *Requirements Engineering* 13, 3 (2008).

[24] Marie Caroline Oetzel and Sarah Spiekermann. 2014. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems* (2014).

[25] I. Oliver. 2016. Experiences in the Development and Usage of a Privacy Requirements Framework. In *2016 IEEE 24th International Requirements Engineering Conference (RE)*.

[26] B. Otto, J. Jürjens, J. Schon, S. Auer, N. Menz, S. Wenzel, and J. Cirullies. 2016. *Industrial Data Space: Digital Sovereignity over Data*. Technical Report. Fraunhofer.

[27] Siani Pearson. 2009. Taking Account of Privacy when Designing Cloud Computing Services. In *Proceedings of the 2009 ICSE Workshop on Software Engineering Challenges of Cloud Computing (CLOUD '09)*. IEEE Computer Society.

[28] Thomas Ruhroth and Jan Jürjens. 2012. Supporting Security Assurance in the Context of Evolution: Modular Modeling and Analysis with UMLsec. In *14th International IEEE Symposium on High-Assurance Systems Engineering, HASE 2012, Omaha, NE, USA, October 25-27, 2012*. 177–184.

[29] K. Schneider, E. Knauss, S. Houmb, S. Islam, and J. Jürjens. 2012. Enhancing Security Requirements Engineering by Organisational Learning. *Requirements Engineering Journal (REJ)* 17, 1 (2012), 35–56.

[30] Sarah Spiekermann. 2012. The Challenges of Privacy by Design. *Commun. ACM* 55 (July 2012), 38–40.

[31] Sarah Spiekermann and Lorrie Faith Cranor. 2009. Engineering Privacy. *IEEE Trans. Softw. Eng.* 35, 1 (Jan. 2009).

[32] Harald Störrle. 2017. How Are Conceptual Models Used in Industrial Software Development? A Descriptive Survey. In *Proceedings of the 21st International Conference on Evaluation and Assessment in Software Engineering (EASE'17)*.

[33] The European Parliament and the Council of the Europeam Union. 2016. Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data . *Official Journal of the European Union* (2016).