Supporting Privacy Impact Assessment by Model-Based Privacy Analysis

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

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

Volker Riediger
University of Koblenz Landau, Germany
riediger@uni-koblenz.de

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

ABSTRACT

According to Article 35 of the General Data Protection Regulation (GDPR), data controllers are obligated to conduct a privacy impact assessment (PIA) to ensure the protection of sensitive data. Failure to properly protect sensitive data may affect data subjects negatively, and damage the reputation of data processors. Existing PIA approaches cannot be easily conducted, since they are mainly abstract or imprecise. Moreover, they lack a methodology to conduct the assessment concerning the design of IT systems. We propose a novel methodology to support PIA by performing model-based privacy and security analyses in the early phases of the system development. In our methodology, the design of a system is analyzed and, where necessary, appropriate security and privacy controls are suggested to improve the design. Hence, this methodology facilitates privacy by design as prescribed in Article 25 of the GDPR. We evaluated our methodology based on three industrial case studies and a quality-based comparison to the state of the art.

CCS CONCEPTS

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

KEYWORDS

Privacy impact assessment, Model-based engineering, Privacy, GDPR, Privacy by design

ACM Reference Format:
Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. 2018. Supporting Privacy Impact Assessment by Model-Based Privacy Analysis. 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/10.1145/3167132.3167288

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

ACM ISBN 978-1-4503-5191-1/18/04…$15.00
https://doi.org/10.1145/3167132.3167288

1 INTRODUCTION

Ensuring data protection and privacy has become a major problem for enterprises that require personal data of their customers to perform their IT services. Article 35 of the GDPR prescribes Privacy Impact Assessment (PIA) [31]. A PIA aims to conduct a systematic risk assessment in order to identify privacy threats and impose technical and organizational controls to mitigate those threats.

According to the GDPR, a PIA shall be conducted prior to the processing, in the early phases of development (design phases). This follows the concept of Privacy by Design (PbD) that is stipulated in Article 25 of GDPR. PbD requires that the design of IT systems must be focused or technically adapted—by implementing appropriate controls—to ensure that, by default, the principles relating to the processing of personal data are respected. In fact, PbD encompasses the entire process of PIA namely, identifying the privacy requirements and privacy threats, performing risk analysis, and choosing proper controls. PIA provides a practical method to establish privacy by design.

Despite the political momentum to establish PIAs, and while the governments in Canada, the UK, Australia, and the US conduct PIAs in critical sectors, the PIA adoption in the IT sector is still rare, particularly in Europe [23]. A possible explanation is the earlier lack of legal incentive. In addition, although a set of legal documents such as the UK PIA handbook [10], and the CNIL's methodology [8] describe the process of conducting PIAs, they generally are not suitable to be a process reference model. They describe a set of generic and abstract steps toward PIAs, and most importantly, they do not consider the concrete design of a system to identify specific design flaws and threats.

In this paper, we investigate the following research questions: **RQ1** Given a system model, how can concrete privacy threats be identified? **RQ2** How can a model-based privacy analysis support a PIA? To address the research questions, we propose a model-based methodology to support a PIA. Analyzing the system models allows a system designer to identify concrete privacy threats in the early phases of system development, and to conduct a privacy impact assessment concerning those threats. Moreover, this methodology identifies a set of proper controls to mitigate the privacy risks arisen from identified privacy threats. We apply our methodology to three industrial case studies from the public administration domain and provide a comparative evaluation to existing approaches.

The paper is organized as follows. In Section 2, the necessary background is provided. In Section 3, we explain how PIA can be

supported by model-based privacy analysis. In Section 4, we present our case studies and evaluation. In Section 5, we discuss related work. Finally, in Section 6, we conclude.

2 BACKGROUND

Below, we present the necessary background for this paper.

2.1 Privacy Impact Assessment

Article 35 of the GDPR uses the term *Data Protection Impact Assessment* to prescribe an *"assessment of the impact of the envisaged processing operations on the protection of personal data"*. In this paper, we use the term *Privacy Impact Assessment* instead, for the following rationale: in [23], the authors argue that privacy is a complex term and is used to appoint different interests—from confidentiality and integrity to transparency and anonymity—therefore, privacy extends beyond the notion of data protection. However, they declare that the privacy threats (identified by [28]) can be addressed through the existing data protection regulation, and therefore the two terms can be considered the same.

The PIA methodology provided in this work is based on the PIA guideline [22] from BSI (Federal Office of Information Security in Germany). In [23], this guideline is extended and introduced as a seven step methodology: (**1**) System characterization. (**2**) Specification of privacy targets. (**3**) Evaluation of degree of protection demand for privacy targets. (**4**) Identification of threats. (**5**) Identification of controls. (**6**) Implementation of controls. (**7**) Generation of PIA report.

Concerning these steps (particularly 3, and 4), it is not specified given a concrete system design, how it is possible to: (I) identify the concrete threats and calculate the risks arising from those threats, and (II) propose proper controls to mitigate those risks and improve a concrete system design. Our PIA methodology aims to answer these two questions.

2.2 Model-Based Privacy & Security Analysis

Developing complex systems is particularly challenging when different interdependent, or conflicting concerns must be handled in those systems. Model-based approaches assist the developers to face such challenges. Using models, the complexity of systems can be handled through abstraction. This enables the developers to focus on main concerns such as privacy and security [14]. Different system models—such as informal usage for communication or learning, and formal usage—are widely used in industry, and UML is in fact the leading language in numerous software domains [29].

Our main contribution is to support privacy impact assessment with model-based approaches. We leverage two model-based approaches to analyze system models, namely UMLsec [16, 19, 20, 27], and a model-based privacy analysis introduced in [3, 4]. The former one provides an approach to develop and analyze software systems, where security requirements are expressed within a system model. The latter one verifies if the principles on processing of personal data—stipulated in Article 5 of the GDPR—are supported by a system's design. Both approaches are based on UML profiles. An open source tool called CARiSMA (http://carisma.umlsec.de) supports both approaches to perform security and privacy analysis. Although an analysis may detect potential vulnerabilities and design flaws in a system, none of both approaches provides any mechanism to further evaluate such results to identify which harmful activities or threats may be exploited from those vulnerabilities.

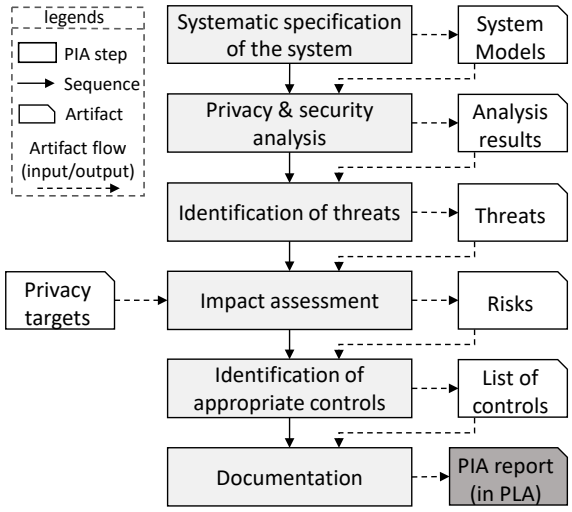

Figure 1: Privacy impact assessment supported by model-based privacy analysis.

3 PRIVACY IMPACT ASSESSMENT BY MODEL-BASED PRIVACY ANALYSIS

Figure 1 demonstrates our PIA methodology. To support *Privacy by Design*, this methodology is to be applied in the early phases of system design. However, it may be also used to conduct a PIA on existing systems. When an existing system is modified, the PIA must be repeated. Iterations of the PIA may be conducted on a system along with the system development or due to system modifications.

3.1 Systematic Specification of the System and its Privacy-Critical Parts

The first step of a PIA is to describe the system. In our PIA methodology, a system model specifies the behavior and the structure of a system. To enable an analysis, this system model is annotated with the privacy and the security profiles (see Section 2.2). The system is modeled using UML. Our methodology is not limited to UML and different modeling languages may be used to model a system. However, this requires appropriate privacy and security profiles.

Moreover, in this step, we verify if conducting a PIA is necessary. According to the GDPR (Article 35, paragraph 3), a PIA shall in particular be required in the case of: (I) a systematic and extensive processing of personal data, (II) processing of special categories of personal data (defined in article 9 of the GDPR), (III) systematic monitoring of a publicly accessible area on a large scale [31]. We leverage the stereotype ≪sensitiveData≫—defined in the privacy profile of [4]—to verify if a PIA must be conducted.

The stereotype ≪sensitiveData≫ specifies that a *NamedElement* in a UML diagram such as an *ObjectNode* is or contains personal data. An ObjectNode in an *activity diagram* annotated with

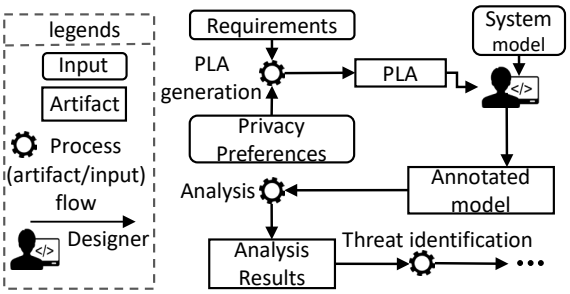

Figure 2: The overview of privacy analysis.

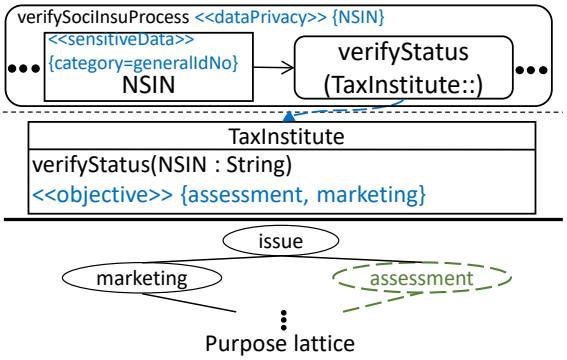

Figure 3: Design model excerpt demonstrating purpose-check [4].

≪sensitiveData≫ specifies that a piece of personal data is processed by performing the business process—represented as the activity diagram—and a PIA must be conducted. However, concerning the three cases in which a PIA is required, a categorization of personal data is essential. ≪sensitiveData≫ currently does not enable any categorization. Therefore, we extend the privacy profile, and define different categories for a piece of personal data.

In the GDPR, different categories of personal data are introduced. (I) Personal data is generally defined in Article 4. (II) Article 9, paragraph 1, refers to special categories of personal data. (III) Article 87 states that specific conditions for the processing of a national identification number or any other identification of general application must be determined. Moreover, in [9], the term *privacy-relevant data* is introduced, which specifies the data that initially are not considered as personal data, however later risks for the privacy of a person or a group or people may become apparent. Therefore we additionally consider *privacy-relevant data* as a category of personal data. This category of data includes the meta data—for instance, system log data such as timestamps that are automatically collected—, and is the super group of all other personal data categories.

We extend the privacy profile and introduce a tag *category* for the stereotype ≪sensitiveData≫. The value of this tag belongs to one of the four categories, namely *commonPersonalData*, *special*, or *generalIdNo*, *privacyRelevantData*. This categorization provides only a baseline for identifying and assessing different categories of personal data, however, concerning different regulations, and specific needs of IT systems, other categories can and should be added.

3.2 Model-Based Analysis

In the second step, the system models from the previous step are analyzed regarding privacy and security requirements. The analysis is performed in a model-based manner based on a set of privacy and security checks. Figure 2 presents the workflow of the tasks performed in this step. Privacy level agreements are used to specify privacy preferences and security requirements. PLAs were originally introduced by the Cloud Security Alliance [5]. A PLA is an appendix to a service level agreement, and provides a structured means to specify privacy preferences, and data security requirements.

In [1], the authors use PLAs to specify the levels of privacy protection that data processors provide—using privacy preferences—and eventually perform a privacy analysis including four privacy checks. Privacy preferences and the privacy checks are based on four key privacy elements, namely *purpose* (authorized reasons to access data), *visibility* (who is authorized to access data), *granularity* (the levels of precision of data), and *retention* (when must a piece of data be deleted or restricted). In our methodology, in addition to the privacy preferences, we specify the security requirements in a PLA. This implies that the analysis (illustrated in Figure 2) involves the UMLsec checks (see Section 2.2), in addition to the four privacy checks. Following an analysis, a set of results is provided, specifying any violations of the security and privacy requirements. In the existing works, no further evaluation of these results is available.

3.3 Identification of the Harmful Activities and Threats

One of the most important steps in privacy impact assessment and basically any known risk assessment methodology, is the identification of threats. In this step, we evaluate the analysis results from the previous step, and identify threats and harmful activities that may exploit the design flaws and violations in the analysis results. We refer to the harmful activities introduced in [28], and the fifteen top cyber-threats and trends introduced in the ENISA threat landscape [13]. In the former one, a complete set of activities that affect privacy and create harm are listed. The latter provides a summary of the most widespread cyber-threats.

We provide a mapping between the harmful activities, the threats, and the checks, showing an excerpt in Table 1. To each harmful activity (16 in total), the relevant checks and threats are mapped. One may use additional scientific sources for the threats and harmful activities, which would then need to be mapped to checks. The output of our analysis is a set of threats and harmful activities resulting from the present design flaws and vulnerabilities.

For instance, consider Figure 3, which demonstrates excerpts of an activity diagram and a class diagram. The activity diagram expresses a business process in which a *National Social Insurance Number (NSIN)* is processed in a public administration and sent to a tax institute for verifying the status of the person to whom the *NSIN* belongs. Furthermore a purpose lattice is provided. Such a lattice specifies a set of all possible purposes including authorized purposes to process a piece of data. The authorized purposes are depicted with dashed-lines and belong to the privacy preferences

Table 1: An excerpt of the mapping between harmful activities, the privacy/security checks, and the threats.

Harmful activities [28]	Relevant check	ENISA 16 threat [13]
Surveillance	Secure links	15. Cyber espionage
Insecurity	Security checks	2. Web-based attacks, 12. Data breaches
Secondary use	Purpose-check, Retention-check	12. Data breaches, 14. Information leakage
Intrusion	Purpose-check (marketing purpose)	6. Phishing, 7. Spam

Table 2: An excerpt of the mapping between the privacy targets and the privacy/security checks.

Privacy target	Appropriate check
P1.1 Ensuring fair and lawful processing by transparency	Existence of PLA
P1.2 Ensuring processing only for legitimate purposes	Purpose-check
P1.4 Ensuring limited processing for specified purposes	Purpose-check
P1.7 Ensuring data quality, accuracy and integrity	Security checks
P1.8 Ensuring limited storage	Retention-check
P1.9 Ensuring the categorization of personal data	(≪sensitiveData≫ {tag})
P4.2 Facilitating the rectification, erasure or blocking of data	Retention-check
P4.3 Facilitating the portability of data	Visibility-check
P5.2 Facilitating the objection to direct marketing activities	Purpose-check (marketing purpose)
P5.3 Facilitating the objection to data-disclosure to others	Visibility-check
P6.3 Ensuring the effectiveness of technical and organizational measures. A 32.1(d)	Currently by existence of PLA

on processing of a piece of personal data. According to Figure 3, *NSIN* is annotated with ≪sensitiveData≫ {category = generalIdNo} specifying that the *NSIN* is an identification of general application. Therefore, a PIA must be conducted.

An analysis concerning the privacy preferences and the security requirements must be performed. The proper check in this case is *purpose-check* (see [4]), which verifies if an objective of an operation in a class that corresponds to an action in an activity diagram conforms to the authorized purposes. According to Figure 3, ≪objective≫ specifies that the *NSIN* is processed for the purposes *assessment* and *marketing*. However according to the dashed sections of the purpose lattice, the *NSIN* is only authorized to be processed for the purpose of *assessment*. This is a violation. Concerning Table 1, this violation in which a piece of personal data is used for unauthorized purposes, may lead to *secondary use*, and *intrusion*. According to [28], the former one refers to the use of data for reasons unrelated to the initial purposes for which the data is collected. The latter one refers to the activities that can disturb one's life, destroy one's solitude, and make one feel uncomfortable.

3.4 Impact Assessment

In this step, we specify the impact of the identified design flaws and threats on a system. To this end, we conduct a risk assessment, in order to identify *what is at risk*. Generally, a set of privacy targets—that need to be respected by system design—is required to enable a PIA and identify what is at risk. In our PIA methodology we use the privacy targets from the BSI methodology [22, 23]. Privacy targets are derived from the legal privacy principles prescribed in data protection regulations [30, 31]. The privacy principles are often abstract. Engineers think of data quality in more concrete terms [26]. Therefore, the privacy principles must be translated into concrete, and functionally enforceable privacy targets [24, 25].

The proposed privacy targets in the BSI methodology provide only a baseline, and more targets can be added [23]. Concerning the GDPR, we introduce two new privacy targets. We require to specifically mention the categorization of a piece of data as a privacy target (see *P1.9* Table 2).

Moreover, according to Article 32, paragraph 1.(d) of GDPR: "[...]a process for regularly testing, assessing and evaluating the effectiveness of technical and organizational measures for ensuring the security of the processing is needed." Therefore, we introduce *P6.3* to ensure the effectiveness of technical and organizational measures.

To assess the impact of design flaws, the identified threats, and the harmful activities on the privacy targets, we require to extend the mapping provided in Table 1 by the privacy targets. Due to space limitations, we only show an excerpt of the mapping between the privacy targets and the privacy and security checks in Table 2. According to this table, not all the privacy targets are mapped to specific checks. For instance, *P1.1* and *P6.3* are mapped to *Existence of PLA*. In this case, the existence of PLA or specific section in PLA ensure the corresponding privacy targets.

Some of the privacy targets are mapped to one specific privacy check. For instance, *P1.2, P1.4*, and *P5.2* are mapped to the *purpose-check*. In this case, the check is used differently to realize the mapping. For example in case of *P1.2*, we verify if for a piece of personal data, a legitimate purpose(s) in a PLA is specified. For *P1.4*, we verify if a piece of personal data is processed for unauthorized purposes.

For some privacy targets, a new utility of a check is defined to realize the mapping. For instance, *P4.3* refers to Article 20 of GDPR, and implies that a data subject (who provides data) shall always have the access right to his/her personal data. For this, the *visibility-check* is used (see [4]). We verify if an access right to a piece of personal data for a data provider during the whole processing exists.

By evaluating the results of an analysis, and using Table 2, we identify which privacy targets are at risk. For instance, considering the example provided in Figure 3, a *NSIN* is used for unauthorized purposes (particularly the unauthorized purpose *marketing*). Regarding Table 2, the targets *P1.4* and *P5.2* are at risk. After identifying what is at risk, we need to perform a risk assessment. Generally, risk assessment can be performed differently, provided that a likelihood and a severity are obtained for each risk [8].

In privacy domain, calculating the probability that a threat may occur is fuzzy. The authors of [23] discourage the use of threat probability estimation from security risk analysis in the privacy domain, because privacy is related to human emotions [28] and if a human right such as privacy is threatened, the arising risks must be mitigated. The CNIL PIA methodology [8] introduces a scale to estimate likelihoods of threats, and provides a set of variables that affect this scale, such as *opening on the internet, data exchange with third parties, variability of the system*. In our methodology, estimating such likelihoods is not essentially relevant. The systems that we analyze are mostly open on internet, include data exchange with third parties, and are interconnected with other systems. Thus, the CNIL likelihood variables cannot be used to estimate the likelihoods.

In our PIA methodology, similar to the BSI PIA methodology, we do not consider the specific likelihoods, and if a privacy threat exists, we control it. Particularly, our risk assessment is only based on the severities of design flaws and the threats (on privacy targets). If a privacy target is at risk, we predict the potential impact of this risk. This depends on two factors: (I) what kind of personal data are analyzed? And (II) what kind of system is analyzed?

We introduced a categorization for ≪sensitiveData≫ in Section 3.1. This categorization affects our severity estimations. For instance, consider the two cases: (I) a privacy target is at risk, since a piece of personal data from the *special* category is used for unauthorized purposes. (II) A privacy target is at risk, since a piece of personal data from the *commonPersonalData* category is used for unauthorized purposes. The severity in case (I) is higher than in case (II). To estimate the effect of different personal data categories on the estimation of severities, we introduce *Personal Data Category Value (PDCV)*. A PDCV is used to evaluate the criticality of a piece of data from a specific personal data category, in a processing. In Table 3, three different values are assigned to the four categories. The criticality of the two categories *special* and *generalIdNo* are equal and evaluated to the highest value.

Moreover, different perspectives of different stakeholders in different systems may affect the severities. Thus, we need to evaluate the degree of protection demand for each privacy target that is at risk, from a perspective of a stakeholder. We focus particularly on two stakeholders, namely a data subject that provides the data (data controller), and the data processer that performs the processing of personal data. To estimate the degree of protection demand for each privacy target, following an analysis, we generate several questions based on the privacy targets and ask for the feedback of data subjects and data processors. According to Table 3, we introduce the *Impact Value (IV)* to evaluate their feedback. In fact, two impact values are calculated regarding the answers that each stakeholder provides, namely *Data Controller IV (DC-IV)*, *Data Processor IV (DP-IV)*.

Table 3: Personal Data Category Values (PDCVs), and Impact Values (IVs).

PDCV	**Personal data category**
0.25	privacyRelevantData
0.5	commonPersonalData
1	special, generalIdNo
IV	**Impacts**
1-Negligible	Either not affected or may encounter a few inconveniences
2-Limited	May encounter significant inconveniences (may be able to overcome)
3-Significant	May encounter significant consequences (may be able to overcome by difficulties)
4-Maximum	May encounter irreversible consequences (may not overcome)

Table 4: The categorization of the IA Scores.

IA Score range	**Category**
$IA < 4$	Low
$4 \leq IA < 8$	Medium
$8 \leq IA < 12$	High
$12 \leq IA$	Very High

A final score of an *Impact Assessment (IA)* for each privacy target that is at risk is calculated by multiplying all these three values:

$$IA = PDCV \times DC\text{-}IV \times DP\text{-}IV$$

The definition of PDCV, different IVs, their respective values, and the categorization of the IA scores are based on the definitions and the categorizations that are provided in [8, 11].

The analysis of the system model provided in Figure 3 identified that two privacy targets *P1.4*, and *P5.2* are at risk. Two questions must be generated and asked for the feedback of the stakeholders. For instance, concerning P5.2, the question "What would happen when facilitating the objection to direct marketing activities is at risk?" will be generated. Together with this question the concrete detailed information "NSIN is used for the purpose of marketing" will be also generated. Assume that the two stakeholders evaluate the impacts as following: DC-IV is limited, and DP-IV is Maximum. Considering the fact that NSIN belongs to the category *generalIdNo*, the impact assessment score for this privacy target equals to eight ($1 \times 2 \times 4 = 8$).

Eventually, after calculating the IA scores for all the privacy targets that are at risk, the identified risks must be mitigated. The mitigation is performed by the controls that are introduced in the next step. To allow this mitigation and choosing the proper controls, we first need to categorize the IA scores. In Table 4, a categorization of different ranges of IA score into four categories, namely *low*, *medium*, *high*, and *very high* is provided.

3.5 Identification of Appropriate Controls

An important step in a PIA methodology is to identify and recommend appropriate privacy and security controls to mitigate the

risks, and improve the system design. In our PIA methodology, we provide a catalog of privacy and security controls. This catalog is based on the security controls of ISO 27001 [18], the privacy control catalog of NIST [21], the measure catalog of the German IT baseline protection [7], and the privacy strategies that are provided in [15]. Currently our control catalog contains 212 controls.

In order to identify an appropriate set of controls, we extend Table 2 by mapping the controls to the corresponding privacy targets. Due to space limitations, we do not present the list of identified controls in this paper. Following the identification of the privacy targets that are at risk, and regarding the provided mapping, we identify a set of controls that potentially mitigate the risks.

The controls are divided into technical and organizational controls. The technical controls explicitly specify which mechanisms must be incorporated into the system design, in order to mitigate the identified risks. An encryption algorithm or an access control are two examples for technical controls. Organizational controls are mainly management or administrative recommendations. For instance, if a privacy target is at risk since no authorized purposes are specified for a piece of personal data in a relevant section of a PLA, an organizational control recommends to conclude a proper PLA with the purpose specification for personal data.

Furthermore, since the costs of incorporating the controls in a system may vary, we introduce four categories of controls, namely *sufficient*, *medium*, *strong*, and *very strong*. This categorization concerns the categorization of the different ranges of IA (Impact Assessment) scores. If the IA score for a privacy target is *very high*, a control from the category *very strong* is suitable to mitigate the identified risk.

3.6 PIA report

Identifying a common mechanism to report a PIA is rather challenging. In [34], the authors state that it is difficult to find published examples of PIA reports. In their work, they analyze a number of existing published PIAs, and propose criteria to assess the effectiveness of a PIA report. We propose to use Privacy Level Agreements (PLAs) to include PIA reports. As we mentioned before, a PLA aims to specify privacy levels that must be respected by a data processor.

In a PLA several sections specify different aspects such as generic information on a data processor, privacy preferences, and security requirements [5]. We compare the current structure of a PLA proposed by Cloud Security Alliance (https://cloudsecurityalliance.org/) with the criteria introduced in [34], and indicate how a PLA must be extended.

(I) A PIA report must specify if a PIA is performed in the early phases of a system development. Since, in our proposed methodology, the assessment starts in the design phase, it is ensured that we start in the early phases. (II) A PIA report must specify who conducted a PIA. We include such information in the generic section of a PLA. (III) In a PIA report any relevant information on the processing of a piece of personal data must be specified. In a PLA, leveraging ≪sensitiveData≫ and regarding the corresponding activity diagrams, we specify the personal data that is processed, together with relevant information including purpose, visibility, retention, and granularity. (IV) In a PIA report the process of an assessment must be described. In a PLA, according to the six steps of our methodology and the output of each step, we specify several sections to document the artifacts including threats, harmful activities, risk assessments, and proposed controls. The current structure of a PLA already includes a section on security and privacy controls, however, concerning a PIA assessment, the identified controls must be mapped to concrete system design flaws and threats. (V) A PIA report must be published. Since a PLA that is concluded between two parties must be always available, this criterion is fulfilled.

4 VALIDATION

In this section, we first introduce our case studies and discuss the research questions. Afterwards we provide a comparative evaluation to three existing PIA methodologies.

4.1 Case Studies

To evaluate our methodology, we used three industrial case studies from the VisiOn EU project (http://www.visioneuproject.eu/) in which a platform to analyze privacy-critical systems and generate and enforce privacy agreements on the use of personal data, is developed. These case studies belong to privacy-critical domains and model three different Public Administration (*PA*) systems, in which protecting the privacy of personal data is obligatory. In cooperation with the practitioners of each PA system, three system models based on UML were annotated using the privacy and security profiles.

Table 5 provides information on the size of the system models and number of annotations. The first system provides different online administrative services to customers in Municipality of Athens (MoA), a PA in Athens. The example provided in Figure 3 is derived from this case study. In the second case study, the structure and the business processes of a hospital (in Rome, Italy) are modeled. This system uses the personal data of the patients to provide different services to the patients and medical staff. In the third case study a PA system from economic development domain (Ministry of Economic Development in Italy) is modeled. This system performs different tasks such as verifying solvency or generating the tax declaration by processing personal data.

We used the CARiSMA tool (http://carisma.umlsec.de, see Section 2.2) to perform privacy and security analyses (see Figure 2) on the three system models [3]. CARiSMA enables different privacy and security checks. After analyzing the system models in each case study, using Tables 1 and 2, we identified the design flaws and respective harmful activities as well as the threats.

To calculate the IA scores we obtained the feedback of the stakeholders, namely the practitioners of each PA system and the customers, on the privacy targets that were at risk. In each case study, we asked a system expert of the respective PA system directly to provide his/her feedback (*data processor impact value*). *Data controller impact values* (customers' feedback) were calculated through a set of questionnaires. The privacy platform developed in the VisiOn EU project includes different components. One of the components provides a set of privacy questionnaires to obtain the privacy and security needs of the customers. We integrated the questions on the privacy targets into these questionnaires. In each case study, between five and ten customers were questioned through the respective PA. Eventually, the IA scores were categorized, and in each

Table 5: Information on the three case studies.

Domain	UML elements	Annotations
Urban administration	141	33
Public health	167	31
Economic development	309	57

Table 6: The enhanced support of our PIA methodology compared other three PIA methodologies.

	Quality criteria for a best practice PIA process	PIA
1	Early start	–
2	General description of the project	✗
	Information flows	✓
	(Other) privacy implications	–
3	Stakeholder's consultation	–
4	Risk assessment	✓
	Risk mitigation	✓
5	Legal compliance check	–
6	Recommendations and action plan decision	✓
	Implementation of recommendations	✓
	PIA report	–
7	Audit and review	–

✓: enhanced support compared to the 3 approaches
✗: not supported
– : similar to other three approaches

case study using our control list, we were able to recommend a set of controls to mitigate the identified risks.

We successfully applied our PIA methodology to three rather complex industrial software systems. Concerning the explored research questions: **RQ1**: Using model-based privacy and security analyses, we described how concrete design flaws, known harmful activities, and threats may be identified given a system design. Two model-based approaches, supported by a tool (CARiSMA), perform the security and privacy analyses to identify the respective harmful activities and threats. **RQ2**: We introduced a six-step methodology to conduct a PIA. Concerning the case studies, we observed that our proposed PIA methodology supports any PIA methodology by analyzing a system design in the early phases of development. Analyzing a system model does not guarantee that all threats will be identified and the real implementation (code) of a system supports all privacy and security requirements. However, identifying the design flaws in the early phases assists the system implementers to integrate recommended privacy and security controls into the system implementation, and facilitates privacy by design. After performing the PIA in each case study, we gathered the customers' feedback on our PIA methodology, we noticed that the customers' awareness of privacy—particularly the importance of supporting the four privacy elements in the system design—is increased.

4.2 Comparative Evaluation

In order to validate the effectiveness of the proposed PIA methodology, and to verify how this methodology may support existing PIA methodologies, we compare this methodology to three recognized PIA methodologies in Europe, namely the UK PIA code of practice [10], the CNIL PIA tools and templates [8], and the BSI PIA methodology [22], according to the seven PIA quality criteria published in [32]. Table 6 demonstrates this comparison.

1. Early start: All four methodologies are used to conduct a PIA from the early phases of a system development.

2. Project description: We do not require to explicitly describe the context of a project such as organizational goals. The three PIA methodologies (UK, CNIL, and BSI) require a general description of the project. Regarding *information flows*, in our proposed PIA, using system models (activity diagrams) we specify how a piece of personal data is processed. The UK PIA code of practice, records the information flows in whichever format—textual descriptions, or models such as flowcharts—however, such models are not technically analyzed. The CNIL and the BSI methodologies describe the processes (information flows) only generically.

3. Stakeholder consultation: All four PIA methodologies support Stakeholder consultation. Although we do not require a general description, we support the perspectives of the stakeholders during impact assessment.

4. Risk management: The UK PIA code of practice only provides a set of generic risks. In CNIL and BSI different templates and guidelines are provided to perform a risk assessment and mitigate the identified risks. However, these templates and guidelines are rather imprecise and abstract. We conduct a risk assessment regarding the identified design flaws, threats, and the categorization of personal data.

5. Legal compliance check: All four PIA methodologies complies with legal requirements and principles. In our work, we updated the list of privacy targets, regarding the privacy principles in the GDPR, and added two new targets.

6. Recommendation and report: All four methodologies provide recommendations to adapt and improve their systems. Chapter 7 of the UK PIA code of practice, the BSI IT baseline protection [7], and the chapter two of CNIL PIA methodology, provide a set of privacy controls to mitigate the identified risks. Similar to these controls we also provide a list of privacy controls. We choose the proper controls according to the conducted risk assessment. Furthermore, we use privacy level agreements to document a PIA report, and generate a structured document to include the results of each step of our PIA methodology.

7. Audit and review: This criterion requires that a PIA report must be externally audited. By documenting a PIA report in a PLA, a formal description of a PIA report is generated, which facilitate the external audit of a PIA report.

The results of the comparison include that although the UK PIA code of practice, the CNIL PIA and the BSI PIA methodologies support the seven quality criteria, they are rather abstract and generic. They do not perform a system level privacy analysis. Moreover, our PIA methodology is supported by a tool (CARiSMA) to perform an analysis.

5 RELATED WORK

The approach described in this paper is a novel methodology to support a PIA by model-based approaches. Our work is motivated

specifically by Article 35 of General Data Protection Regulation (GDPR).

In [23], a Systematic methodology for privacy impact assessment by formally representing a structure to analyze privacy requirements, and assisting practitioners to handle the complexity of privacy regulations, is provided. In [6], a process for data protection impact assessment under European general data protection regulation is provided. In [33], the authors review the existing PIA methodologies, conduct a survey on PIA in the EU, and recommend an optimized PIA framework to European Commission. However in these works, they do not analyze a concrete system design to identify concrete design flaws.

In [12], the authors provide a framework to model privacy threats. However, they do not analyze a concrete system model to identify concrete threats.

[8, 10, 22] provide methodologies and best practices to conduct a PIA in the UK, France, and Germany. Moreover, the ISO 27000 family of standards on information security management [18], and the ISO 31000 risk management standard [17] are recognized standards to keep information assets secure, and generally manage risks in organizations. However, these legal methodologies and standards are rather abstract and imprecise, and cannot be used as a concrete methodology to conduct a PIA.

6 CONCLUSION

We introduced a novel methodology to support privacy impact assessment using model-based privacy and security analyses. Complying with the GDPR, we introduced a categorization of personal data. To fully support the privacy principles that are prescribed in the GDPR, we introduced two new privacy targets. Moreover, we presented a mechanism to calculate the impact of the threats on the privacy targets. We applied our methodology to industrial scenarios.

In our ongoing work, we investigate how we may extend the existing model-based privacy analysis [4] in compliance with the Article 5 of the GDPR, where *purpose* is the fundamental privacy element, to support privacy protection in a specific industrial ecosystem where a network for trusted data exchange between enterprises is established [2]. In the future, we will investigate how recommended controls may be applied to a system design to ensure their effectiveness.

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).

REFERENCES

[1] A. S. Ahmadian and J. Jürjens. 2016. Supporting Model-Based Privacy Analysis by Exploiting Privacy Level Agreements. In *2016 IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*.

[2] 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 *Proceedings of ACM SAC Conference (SAC 18)*. ACM, New York, NY, USA. accepted.

[3] 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

[4] Amir Shayan Ahmadian, Daniel Strüber, Volker Riediger, and Jan Jürjens. 2017. Model-Based Privacy Analysis in Industrial Ecosystems. In *Modelling Foundations and Applications*. Springer International Publishing.

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

[6] F. Bieker, M. Friedewald, M. Hansen, H. Obersteller, and M. Rost. 2016. *A Process for Data Protection Impact Assessment Under the European General Data Protection Regulation.* 21–37.

[7] Bundesamt für Sicherheit in der Informationstechnik. 2016. BSI-Grundschutz Katalog. (2016).

[8] Commission nationale de l'informatique et des libertés. 2015. Privacy Impact Assessment - Tools. (2015).

[9] G. Danezis, J.D. Ferrer, M. Hansen, J.H. Hoepman, D. Le Métayer, R. Tirtea, and S. Schiffner. 2015. Privacy and Data Protection by Design - from policy to engineering. (2015).

[10] Data Protection Act. 2014. *Conducting privacy impact assessments code of practice.* Technical Report. Information Commissioners Office (ICO).

[11] Data Protection Authorities of Greece and Germany, Clara Galan Manso, and Slawomir Gorniak. 2013. *Recommendations for a methodology of the assessment of severity of personal data breaches.* Technical Report. The European Union Agency for Network and Information Security.

[12] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. 2011. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering* 16 (2011).

[13] ENISA. 2016. *ENISA Threat Landscape Report 2016.* Technical Report. The European Union Agency for Network and Information Security.

[14] R. France and B. Rumpe. 2007. Model-driven Development of Complex Software: A Research Roadmap. In *2007 Future of Software Engineering (FOSE 07)*.

[15] J. Hoepman. 2014. *Privacy Design Strategies.* Springer, 446–459.

[16] 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*. Idea Group, 190–219.

[17] International Organization for Standardization. 2009. *ISO/IEC 31000 Risk management - Principles and guidelines.* Technical Report.

[18] International Organization for Standardization. 2013. *ISO/IEC 27001 Information technology - Security techniques -Information security risk management.* Technical Report.

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

[20] 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.

[21] National Institute of Standards and Technology. 2013. Security and Privacy Controls for Federal Information Systems and Organization. (2013).

[22] M.C. Oetzel, S. Spiekermann, I. Grüning, H. Kelter, and S. Mull. 2011. *Privacy Impact Assessment Guideline.* Technical Report. Bundesamt für Sicherheit in der Informationstechnik.

[23] Marie Caroline Oetzel and Sarah Spiekermann. 2014. A systematic methodology for privacy impact assessments: a design science approach. *European Journal of Information Systems* (2014), 126–150.

[24] Martin Rost. 2011. Datenschutz in 3D - Daten, Prozesse und Schutzziele in einem Modell. *Datenschutz und Datensicherheit* 35, 5 (2011), 351–354.

[25] Martin Rost and Andreas Pfitzmann. 2009. Datenschutz-Schutzziele - revisited. *Datenschutz und Datensicherheit* 33, 6 (2009), 353–358.

[26] Monica Scannapieco, Paolo Missier, and Carlo Batini. 2005. Data Quality at a Glance. *Datenbank-Spektrum* 14 (2005), 6–14.

[27] 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.

[28] Daniel J. Solove. 2006. A Taxonomy of Privacy. *University of Pennsylvania Law Review* 154 (2006).

[29] 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)*.

[30] The European Parliament and the Council of the Europeam Union. 1995. Directive 95/46/EC. *Official Journal of the European Union* (1995).

[31] The European Parliament and the Council of the Europeam Union. 2016. Regulation (EU) 2016/679. *Official Journal of the European Union* (2016).

[32] David Wright and Paul De Hert. 2012. *Privacy Impact Assessment.* Springer Netherlands.

[33] David Wright and Kush Wadhwa. 2013. Introducing a privacy impact assessment policy in the EU member states. *International Data Privacy Law* 3 (2013), 13.

[34] David Wright, Kush Wadhwa, Paul De Hert, and Dariusz Kloza. 2011. *A Privacy Impact Assessment Framework for Data Protection and Privacy Rights.* Technical Report.