Department of Computer Science

Towards an effective PIA-based Risk Analysis: An Approach for Analysing Potential Privacy Risks

Majed Alshammari and Andrew Simpson

CS-RR-18-01

Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford, OX1 3QD

Towards an effective PIA-based Risk Analysis: An Approach for Analysing Potential Privacy Risks

Majed Alshammari and Andrew Simpson
Department of Computer Science, University of Oxford
Wolfson Building, Parks Road,
Oxford OX1 3QD, UK
Email: firstname.secondname@cs.ox.ac.uk

Abstract—**The use of Privacy Impact Assessments (PIAs) has become common practice in a variety of jurisdictions since the mid 1990s. They play a crucial role in achieving privacy protection for data subjects and in supporting risk management for organisations. Many guidance documents have been published to help support organisations in performing PIAs and in achieving their intended benefits. However, these documents vary noticeably in their comprehensiveness and quality. From an engineering perspective, the core of a PIA is a risk assessment, which typically follows a step-by-step process of risk identification and risk mitigation. In order for a PIA to be holistic and effective, it needs to be complemented by an appropriate privacy risk model that considers legal, organisational, social and technical aspects. We propose a methodical approach for identifying and analysing potential privacy risks. It is built upon a conceptual model that represents the main factors that have impacts on privacy risks along with their meanings, properties and relationships. Then, we illustrate its use in the analysis of eToll pricing systems. We argue that this contribution lays the foundation for developing systematic and rigorous PIA methodologies.**

I. Introduction

In practice, processing personal data inappropriately may lead to privacy violations or harms. To avoid such violations and harms, there is an increased recognition that the potential impacts of processing operations need to be proactively assessed in the early stages of the design process [1]. This has led to the emergence of the concept of a Privacy Impact Assessment (PIA). A PIA is defined as a process that identifies and mitigates the impact of an initiative on privacy with stakeholders' consultation [2]. PIAs are now mandated by, for example, the EU General Data Protection Regulation (GDPR) [3]. However, they do not precisely illustrate how a risk assessment of a PIA should be performed [1]. A PIA tends to focus more on legal and organisational aspects than technical details [1]. In order for a PIA to be holistic and effective, it is necessary for it to be complemented by an appropriate privacy risk model that considers legal, organisational, social and technical aspects.

In systems engineering, risk assessment goes beyond identifying technical risks of the system being developed; however, this requires a better understanding of social perceptions and reasonable expectations that are derived from social norms [4] and [5]. Privacy is not strictly a technical concept; it is a multifaceted concept that requires multidisciplinary considerations [6]. Privacy engineering, therefore, requires a sufficiently robust privacy risk model to identify potential privacy risks. The identified risks can then be addressed through risk management approaches, which include the selection and application of risk controls. In this report, we extend prior work in this research area by referring to fundamentals from the broader literature to underpin the main concepts of PIAs along with their meanings and properties. In addition, we show how these concepts relate to one another and illustrate their use in the analysis of eToll pricing systems. In addition, we present a methodical approach that illustrates the main steps of identifying and analysing potential privacy risks in a meaningful manner. We argue that this contribution lays the foundation for systematic and rigorous PIA methodologies.

This report is organised as follows. Section II gives an overview of existing PIA processes and illustrates the main characteristics of PIAs. Section III explains and discusses the shortcomings of existing PIA processes. In addition, it presents a set of criteria against which systematic and rigorous PIA methodologies may be measured. Section IV presents the main concepts of privacy risk analysis, their properties and relationships. Section V presents a methodical approach that illustrates the main steps of identifying and analysing privacy risks in a meaningful manner. Section VI introduces the European Electronic Toll Service (EETS), the aim of which is to support interoperability between electronic road toll systems, and which we shall use as an illustrative case study. Finally, Section VII summarises the main contributions, and outlines our plans for future work in this area.

II. Background and Motivation

Ensuring that the processing of personal data is conducted fairly and lawfully is one of the main challenges in the context of data protection. This challenge has raised concerns over data-processing activities that may lead to privacy violations or harms. *Privacy by Design (PbD)* [7] has been advocated as a response to these challenges — mainly to meet legal obligations, mitigate potential privacy risks, achieve accountability and enhance user trust [6]. PbD is a strategic management and engineering approach that minimises, mitigates or eliminates potential privacy risks by applying administrative and technical measures [8]. 'Proactive not Reactive; Preventative not Remedial' is a principle of PbD, which emphasises that the PbD approach is characterised by proactive measures rather than reactive ones [9]. As such, the PbD approach anticipates and prevents privacy-related adverse events; i.e. it does not wait for privacy threats to materialise, nor does it provide remedies for resolving negative consequences of such events once they have occurred [9].

The integration of such a proactive approach to privacy into risk-management processes will be of great benefit to organisations. It needs to complement risk-management frameworks to simplify their implementation [10]. To ensure that privacy risks are successfully mitigated, it is important to make certain that the principles of PbD are not only embedded in an organisation's risk-management process, but also in all of its processes dealing with personal data, such as the software development process [10]. To realise the concept of PbD in the system development lifecycle (SDLC), potential privacy risks need to be proactively analysed and their potential harms need to be appropriately assessed [11]. In some jurisdictions, 'legal compliance checks' [11] or 'prior checking' [12] are the most commonly used privacy assessment procedures. These procedures are often not conducted by engineers; rather, auditors, lawyers or data protection authorities go through a check-list to check compliance with legal frameworks [11]. With the advent of information and communication technologies, holistic and effective impact assessments are considered as complements to, or replacements for, these assessment procedures [11]. This has contributed to the emergence of the concept of *a Privacy Impact Assessment (PIA)*.

The term PIA has been associated with a variety of meanings over time and across jurisdictions. Clarke [12] defines a PIA as a systematic process for assessing the potential impacts on privacy of an initiative. Similarly, Wright [2] defines a PIA as a methodology for assessing the impacts on privacy of an initiative with stakeholders' participation to avoid or at least mitigate their negative impacts or consequences. From an engineering perspective, Oetzel and Spiekermann [11] define a PIA as a risk-assessment methodology that is proactively used in the design phase of a given software system to be privacy-friendly and compliant with legal frameworks and standards. Thus, a PIA is more than a tool: it is an ongoing process that begins at the earliest possible stages [13]. As such, PIAs are considered as a key means to address one of the main concerns of embedding privacy into the early stages of the design process, which is the manifestation of PbD [14]. Existing PIA processes strive to achieve the aim of PbD by applying its foundational principles [11].

However, a PIA needs to be distinguished from other procedures undertaken by organisations, such as privacy issues analysis, legal compliance checks and privacy review or audit [12], [15] by the following characteristics, as discussed in [12].

1) It is conducted for a particular initiative — i.e. it is distinct from a privacy strategy formulation that is conducted from a corporate perspective.
2) It is anticipatory in nature — i.e. it is distinct from a privacy audit, which is an assessment conducted after an initiative is completed.
3) It has broad scope with respect to privacy dimensions — i.e. it is distinct from other limited scope processes, such as *a data privacy impact assessment*, which assesses the impacts of an initiative on the privacy of personal data.
4) It has broad scope with respect to participating stakeholders — i.e. it adopts a multi-perspective approach that takes into account privacy concerns of multiple stakeholders, including data subjects. It is distinct from an internal cost-benefit analysis or internal risk assessment.
5) It has broad scope with respect to multiple stakeholders' needs, expectations and concerns — i.e. it is distinct from legal compliance checks.
6) It is oriented towards anticipating and mitigating potential privacy risks — i.e. it is distinct from a preliminary privacy issues analysis.
7) It is an assessment process that should begin at the earliest possible stages. It is not only an impact statement that identifies potential consequences of a current or proposed initiative.
8) It necessitates intellectual engagement — it is not only a check-list.

Crucially, existing PIA guidance documents refer to a PIA as a part of risk management — as it identifies, analyses, assesses and mitigates potential privacy risks [2]. The core of a PIA is a risk assessment, which typically follows a step-by-step process of both risk identification and risk mitigation [11]. Although PIAs are expected to follow the same philosophy, existing PIA processes largely fall short with respect to this [11]. These limitations leave a number of open questions that need to be answered:

1) How can we develop a privacy risk model that defines and/or refines key concepts and assessable risk factors, as well as the relationships among the factors?
2) How can we identify potential privacy risks in a contextual and comprehensive manner to ensure the provision of end-to-end privacy protection?
3) What is the appropriate level of detail for such a model?

These questions motivate our contribution towards a step-by-step process that identifies and analyses potential privacy risks in a contextual, comprehensive and concrete manner.

III. An Analysis of PIA processes

In this section, we discuss the shortcomings of existing PIA processes with regards to the development of an appropriate privacy risk model. Then, we establish a set of success criteria for developing a methodical approach that helps identify and analyse potential privacy risks in a meaningful manner. Such a method lays the foundation for systematic and rigorous PIA methodologies.

A. A critique of existing PIA processes

Even though the guidance documents that have been published within several jurisdictions have useful elements and benefits, they also have some shortcomings with regards to their underlying processes [15].

Existing PIA processes cannot be applied easily — not least because they are imprecise, lengthy or improperly structured [11]. Further, they do not typically support the integration of a PIA into a risk-management process [13]. In addition, there is a lack of proper guidelines and conceptual models that sufficiently support privacy risk assessments [11]. Even though PIAs have been mandated in some jurisdictions, there is a lack of standards that illustrate how these PIAs can be conducted systematically [11]. This criticism can be decomposed into a number of concrete limitations.

1) Insufficient representation of data-processing activities: In order to identify data-processing activities that may lead to privacy violations or harms, it is essential to represent these activities in a way that is amenable to risk analysis and compliance checking. We acknowledge the importance of describing systems in multiple views as proposed in [11]. With a focus on a data view, however, Data Flow Diagrams (DFDs) alone may not be sufficient in providing the appropriate level of detail that is needed for risk assessment. As such, rigorous data models need to be adopted to support the management and traceability of the processing and flow of personal data, as well as to help support identifying the planned, actual and potential data flows and processing. In particular, such data models are expected to represent data-processing activities in a comprehensive manner and at an appropriate level of abstraction. This includes: personal data items, data-processing activities, processing purposes, involved actors, and their roles and responsibilities. Such information helps support establishing the context in which personal data is processed, and identifying the boundaries of the system in question.

2) Inappropriate translation of abstract privacy principles: In a risk assessment, it is essential for engineers to understand what to protect and by which means. Some PIA processes, such as the BSI IT-Grundschutz [16], apply security risk analysis to privacy principles, which are typically given at a high level of abstraction, instead of relying upon a set of concrete protection goals. This, in turn, reduces privacy protection to the concepts of anonymity, pseudonymity, unobservability and unlinkability [17], [11]. Thus, targets of evaluation — i.e. personal data, data-processing activities, along with associated constrains — need to comply with legal frameworks and standards, and ensure that they will not lead to potential privacy violations and harms. These targets define the scope of PIAs. Abstract privacy principles are semantically different from concrete data-processing activities; therefore, it is difficult to use them for assessing these activities and describing design decisions at an architectural level. Accordingly, privacy principles need to be translated into concrete and auditable protection goals to aid engineers in specifying design objectives.

3) Incomplete model of privacy risk factors: In order to conduct an appropriate privacy risk analysis that goes beyond a traditional security analysis, it is essential to develop a risk model that defines the key risk factors that have an impact on privacy risks, such as risk sources, privacy weaknesses, feared events and privacy harms, as well as to establish a conceptual relationship among these factors [1]. In risk assessments, risk factors are defined as characteristics used in risk models as inputs to estimate levels of risks in particular contexts. Existing PIA guidance documents, however, are not accompanied with proper guidelines or conceptual models that describe key risk factors to sufficiently support privacy risk assessment [11].

B. Beyond the critique

From a technical perspective, PIAs need to be complemented by an appropriate privacy risk model that goes beyond traditional security risk models. Such a model needs to consider not only legal and organisational aspects, but also societal and technical aspects. The model needs to refer to fundamentals from the legal privacy literature to underpin the main concepts, the key risk factors and the conceptual relationship between these factors. Such a model addresses the first question of Section II ("How can we develop a privacy risk model that defines and/or refines key concepts and assessable risk factors, as well as the relationships among the factors?") by providing a complete privacy risk model that can be used to determine the degree to which privacy is required.

Most importantly, a privacy risk model needs to adopt a sufficiently robust model that facilitates end-to-end privacy protection and serves as the basis for the identification, analysis and assessment of potential privacy risks in a proactive, comprehensive and concrete manner. Such a robust model needs to sufficiently and contextually represent data-processing activities in a way that is amenable to risk analysis and compliance checking. Such a model addresses the second question of Section II ("How can we identify potential privacy risks in a contextual and comprehensive manner to ensure the provision of end-to-end privacy protection?") by providing a sufficient representation of data-processing activities and translating abstract privacy principles into concrete protection goals.

In addition, an appropriate analysis approach needs to be adopted to systematically describe how combinations of risk factors are identified and analysed. Such an approach needs to consider the appropriateness of the starting points of risk assessment and the level of abstraction in the context of privacy and data protection. Such an analysis approach addresses the third question of Section II ("What is the appropriate level of detail for such a model?") by providing a step-by-step process for analysing potential privacy risks at an appropriate level of detail.

C. Success criteria for privacy risk assessment

We now identify a set of success criteria for developing an appropriate privacy risk analysis approach that can be used to complement a PIA to support the implementation of its core activity — i.e. the risk assessment — in a systematic manner.

1) The key risk concepts are defined and/or refined in the context of privacy and data protection.
2) The relevant and assessable risk factors relating to privacy are defined and/or refined at an appropriate level of detail.

3) The conceptual relationships among these factors are clearly defined and/or refined.
4) The analysis approach that describes how combinations of risk factors are identified and analysed is illustrated to ensure analysing potential privacy risks at a consistent level of detail.

IV. A Privacy Risk Model

A number of privacy risk-management processes, frameworks and methodologies have been proposed, such as the Methodology for Privacy Risk Management [18] and the Privacy Risk Management (PRM) [10], which are both based on the ISO 31000 Risk Management Framework [19].

For our purpose, we review two privacy risk analysis methodologies [1], [18] upon which we build by refining the concepts, risk factors and relationships among these factors. We have chosen these models as they define and distinguish the key notions, risk factors and relationships among these factors in the context of privacy and data protection. There is no denying that different models can lead to different levels of detail in characterising key risk factors. To compare, we refer to fundamentals from the legal privacy literature to underpin the key concepts and risk factors along with their meanings, properties and relationships. In particular, we refer to the boundaries of privacy harm [20] to understand the specific characteristics and categories of privacy harms. In addition, we refer to Solove's taxonomy [21] to understand the specific characteristics of adverse privacy events and associated categories. Finally, we leverage the concept of contextual integrity [22] to understand the main characteristics of appropriate flow of personal data with reference to context-relative informational norms, from which vulnerabilities can be derived. In particular, we consider the starting point of the risk assessment and the level of detail as points of reference for our analysis.

A. The key concepts and privacy risk factors

In order to carry out an appropriate privacy risk analysis that goes beyond traditional security analysis, we define and/or refine the basic concepts used in conducting risk assessments in the context of privacy and data protection. We should note that the only risks to consider are those arising from the processing of personal data that have adverse impacts on the privacy of data subjects.

1) Threats: A threat is an event or action with the potential for privacy violation, or to adversely impact the privacy of data subjects through the processing of personal data via inappropriate collection, retention, access, usage, disclosure or destruction. In our risk model, the threat concept is abstractly represented: it can be decomposed into a threat source and a threat event.

- *Threat sources.* A threat source is an entity with capability to process (lawfully or unlawfully, fairly or unfairly) data belonging to a data subject and whose actions may instantly and/or eventually, accidentally or deliberately manifest threats, which may lead to privacy violations or harms. Each type of a threat source can be characterised by a set of attributes: type (insider or outsider; individual, institution or government; human or non-human), motives (stemming from the value of personal data), resources (including skills and background knowledge that helps re-identify data subjects), role (represents the way in which a concerned entity participates in processing operations, such as normal user, privileged user, service provider, etc.), and responsibility.
  The specified attributes of a threat source are used to assess the capability of exploiting vulnerabilities. As such, a threat source is more relevant to vulnerability analysis than impact assessment, i.e. impact is independent of vulnerability and threat analysis — in practice it is irrelevant whether the threat event flows from an internal or external threat source whose actions are accidental or deliberate. In security risk analysis, threat actors, threat sources and risk sources are often referred to as *attackers or adversaries*. In the context of privacy and data protection, however, we use the concept of a threat source to ensure that it can be used appropriately for modelling actors with malicious and benign purposes.
  Joyee De and Le Métayer [1] use the concept of risk source to refer to both unauthorised entities processing personal data and entities with legitimate processing capabilities. In [18], risk sources are those who act, accidentally or deliberately, on the supporting assets, on which the primary assets rely. Accordingly, threat sources who act, accidentally or deliberately, on the primary assets are not modelled. As such, we refine these concepts to be used appropriately in the context of privacy and data protection at an appropriate level of abstraction. With regards to threat sources who act on the supporting assets, we refine the standard definition threat action. *A threat action* is an intentional act (actively or passively) through which a threat source exploit the vulnerabilities of the supporting assets. It is important to separate the concept of the threat action to engage with the supporting asset and the threat event when a threat source acts against the primary asset.
- *Threat events.* A threat event is a technical event that may happen at specific points in time which has an effect, consequence or impact, especially a negative one, on the privacy of data subjects. Such events involve adverse actions justified by reference to personal data — i.e. *what can go wrong*. A threat event is a possible source of privacy violations or harms: it occurs as a result of a successful exploitation of one or more vulnerabilities by one or more threat sources. Each type of threat event can be characterised by a set of attributes: nature (continuous or discrete; excessive or necessary; anticipated or unanticipated), scope (an individual, a specific group of individuals or whole society), and category (according to the taxonomy of privacy).
  Joyee De and Le Métayer's PRIAM [1] and the Methodology for Privacy Risk Management [18] use the concept

of ‘feared events’. By referring to them as feared events, we may limit those to internal and unpleasant emotions and perceptions caused by the threat. As such, we use the notion of ‘threat events’ to describe harmful or unwanted events that may not be anticipated by data subjects. Since these events not only describe the data subject’s perceptions, we prefer to use threat events to describe unwanted, unwarranted or excessive processing activities, which will lead to actual adverse consequences. They refer to a non-exhaustive list of common categories of feared events that an analyst should consider. However, we prefer to consider a well-known classification of such events.

For the purpose of this report, we consider only technical threats that are processing-related, not those caused by natural disasters, power failures, etc. In particular, we focus on data-processing activities, which are composed of adverse actions that are justified by reference to personal data, and events that cause the performance of these actions, that can and do constitute privacy violations or create privacy harms.

2) Privacy vulnerabilities: A vulnerability is a weakness or deficiency in: personal data modelling; the design or implementation of processing operations; or privacy controls, whether these controls are technical, organisational or legal, that makes an exploitation of an asset more likely to succeed by one or more threat sources. Successful exploitations lead to threat events that can result in privacy violations or harms.

In the context of privacy and data protection, assets can be classified into *primary assets* and *supporting assets* [18]. The former refers to personal data that is directly concerned with processing operations, as well as processes required by legal frameworks and standards. The latter refers to system components on which the primary assets rely, such as hardware, software, people, etc. For the purpose of this report, we focus on the primary assets and associated vulnerabilities — i.e. *what we are trying to protect*. Each type of vulnerability can be characterised by a set of attributes: exploitability and severity. These are used to estimate the level of a vulnerability — i.e. its seriousness.

The Methodology for Privacy Risk Management [18] uses the concept of vulnerability, which refers to characteristics of a supporting asset that can be exploited by risk sources and allowing threats to occur. In contrast, Joyee De and Le Métayer [1] use the concept of ‘privacy weakness’ to refer to a weakness in the data protection mechanisms — whether this weakness is legal, technical or organisational. By using this concept, they aim to include weaknesses that may not be considered by using the concept of vulnerability, such as inappropriate functionality from which privacy harms may stem. As such, we use the concept of vulnerability with a broader view to not identify them only within data protection mechanisms. Privacy vulnerabilities can be found in the implemented privacy controls and the specified processing operations along with required personal. In addition, we use the classification of assets of [18].

3) Privacy violations: A privacy violation is an unlawful and/or unfair action that accidentally or deliberately breaches privacy-related laws, regulations, unilateral policies, contracts, cultural norms or standard principles. Such actions are triggered by occurrences of threat events that result from the successful exploitation of one or more vulnerabilities. In reality, inappropriate processing of personal data may lead to privacy violations, which may involve a variety of types of activities that may lead to privacy harms [21]. Most importantly, the presence of a privacy violation does not mean that it will necessarily create actual privacy harm. This indicates that privacy harms can occur without privacy violations and vice versa [20]. Unauthorised access to sensitive personal data without actual adverse action, for example, making a judgement, and which no one ever knows about is an example of a privacy violation without privacy harms. As such, we distinguish between privacy violations and privacy harms. Each type of privacy violation can be characterised by a set of attributes: type (unlawful or unfair), degree (excessive or limited) and scope (an individual, a group of individuals or whole society).

Joyee De and Le Métayer [1] and the Methodology for Privacy Risk Management [18] do not distinguish between privacy violations and harms. It is well understood that these methodologies focus on feared events and their potential impact — i.e. consequences that each feared event may have on the identity and privacy of data subjects and human rights or civil liberties.

4) Privacy harms.: A privacy harm is the adverse impact (incorporeal, financial or physical) of the processing of personal data on the privacy of a data subject, a specific group of data subjects or the society as a whole, resulting from one or more threat events. A widely held view conceptualises a privacy harm as the negative consequence of a privacy violation [20]. However, privacy harms are related to, but distinct from, privacy violations. This implies that it is not necessary for an actor to commit a privacy violation for a privacy harm to occur and vice versa — i.e. a privacy harm can occur in the absence of a human actor [20]. It is important in a privacy risk analysis to choose an abstract definition of privacy harms to ensure that all possible negative impacts are considered. The main sources of such harms include: previous privacy breaches documented or discussed in the literature, case law, recommendations published by Data Protection Authorities (and related or similar organisations) and the points of view of multiple stakeholders.

Each privacy harm can be characterised by a set of attributes: type (subjective or objective), category (incorporeal, financial or physical), adverse consequences (last for a short time, last for a certain length of time or last for a long time), and affected data subjects (a data subject, a specific group of data subjects, or whole society). To distinguish between subjective and objective categories of privacy harm, the former represents the perception of inappropriate processing of personal data that results in unwelcome mental states, such as anxiety, embarrassment or fear, whereas the latter represents the actual adverse consequence, such as identity theft that stems from the potential or actual inappropriate processing of personal data.

The Methodology for Privacy Risk Management [18] uses the concept of prejudicial effect to assess how much damage would be caused by all the potential impacts. As such, feared events are ranked by estimating their severity based on the level of identification of personal data and the prejudicial effect of these potential impacts. To identify potential impacts of feared events, consequences on the identity and privacy of data subjects and human rights or civil liberties need to be identified. This means that it does not characterise privacy harms to facilitate their identification and analysis. In contrast, Joyee De and Le Métayer [1] use the concept of privacy harms with specific attributes and categories. In our approach, we use the same concept with more detailed to identify privacy harms at a detailed level of abstraction according to the properties and boundaries identified in [20].

B. The relationship between the privacy risk factors

Figure 1 illustrates the conceptual relationships among the key risk factors in our privacy risk model. The action of a threat source on the primary assets may happen through one or more threats, which, in turn, may exploit one or more vulnerabilities. The successful exploitation leads to one or more threat events that would result in zero or more privacy violations or privacy harms.

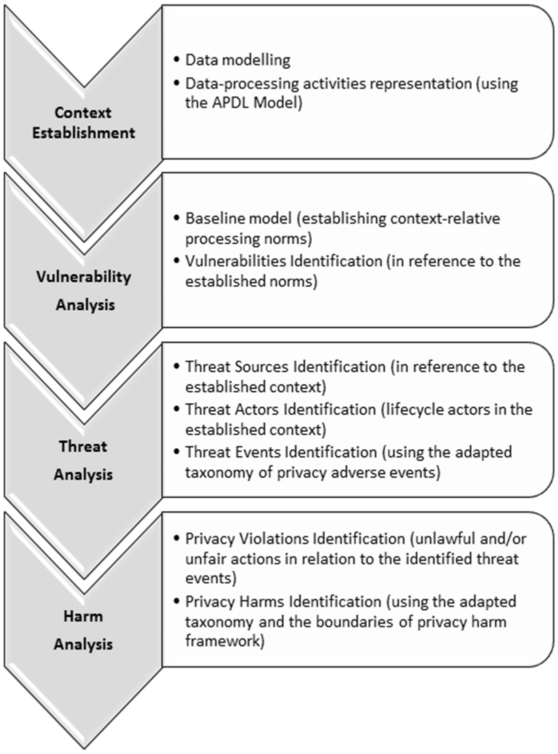
Context Establishment
• Data modelling
• Data-processing activities representation (using the APDL Model)
Vulnerability Analysis
• Baseline model (establishing context-relative processing norms)
• Vulnerabilities Identification (in reference to the established norms)
Threat Analysis
• Threat Sources Identification (in reference to the established context)
• Threat Actors Identification (lifecycle actors in the established context)
• Threat Events Identification (using the adapted taxonomy of privacy adverse events)
Harm Analysis
• Privacy Violations Identification (unlawful and/or unfair actions in relation to the identified threat events)
• Privacy Harms Identification (using the adapted taxonomy and the boundaries of privacy harm framework)

Fig. 2. The main steps of the analysis approach.

V. An Analysis Approach

Risk analysis approaches differ with respect to the starting points of risk assessments and levels of abstraction. In order for risk assessments to be effective, they need to synthesise multiple analysis approaches to identify the key factors of risk. Potential privacy risks need to be identified, analysed and assessed in a systematic manner. As such, our analytical approach consists of four steps. The first step is to establish the context in which personal data is processed. The second step is to identify and analyse all possible vulnerabilities in this particular context. The third step is to identify and analyse potential threat sources and events. The fourth step is to identify and analyse potential privacy violations and harms in this context. Figure 2 illustrates the main steps of the analysis approach. We consider each in turn.

A. Context Establishment

Establishing the context in which personal data is processed plays a crucial role in understanding the scope under consideration by identifying all the useful information for privacy risk analysis. This includes the types of personal data to be processed (primary assets that need to be protected), along with its sources; the purposes for, and the manner in which, this data is processed; involved actors and their assigned roles and responsibilities; relevant legal frameworks and standards; and domain-specific constraints.

As mentioned in Section IV-A2, primary assets are classified into *personal data*, which relates to an identified or identifiable individual, and *processes*, which refer to both actual data-processing activities and privacy-related processes required by legal frameworks and standards. As such, personal data, associated processes and involved actors need to be represented in a way that is amenable to analysis. We acknowledge the importance of describing systems in multiple views as proposed in [11]. However, we emphasise the importance of data-management models that represent data and associated processing activities at a detailed level of abstraction. We believe that data lifecycles are better at describing processing activities in a detailed level of abstraction — i.e. they categorise and represent these activities in relation to the main stages of the lifecycle: from collection to destruction. The Abstract Personal Data Lifecycle (APDL) Model [23] was developed to represent data-processing activities in a way that is amenable to analysis and compliance checking. It represents the personal data lifecycle in terms of lifecycle stages, along with associated activities and involved actors. It can be used to complement a PIA for describing the planned, actual and potential processing of personal data, which, in turn, helps facilitate the management and traceability of the flow of personal data from collection to destruction [23].

Accordingly, we adopt the APDL model to represent the primary assets, along with involved actors. Personal data is represented in the *DataModelling* stage. This stage represents the relevant objects, associated properties, relationships and constraints for the purpose of specifying the minimum amount of required personal data. Processes are abstractly represented

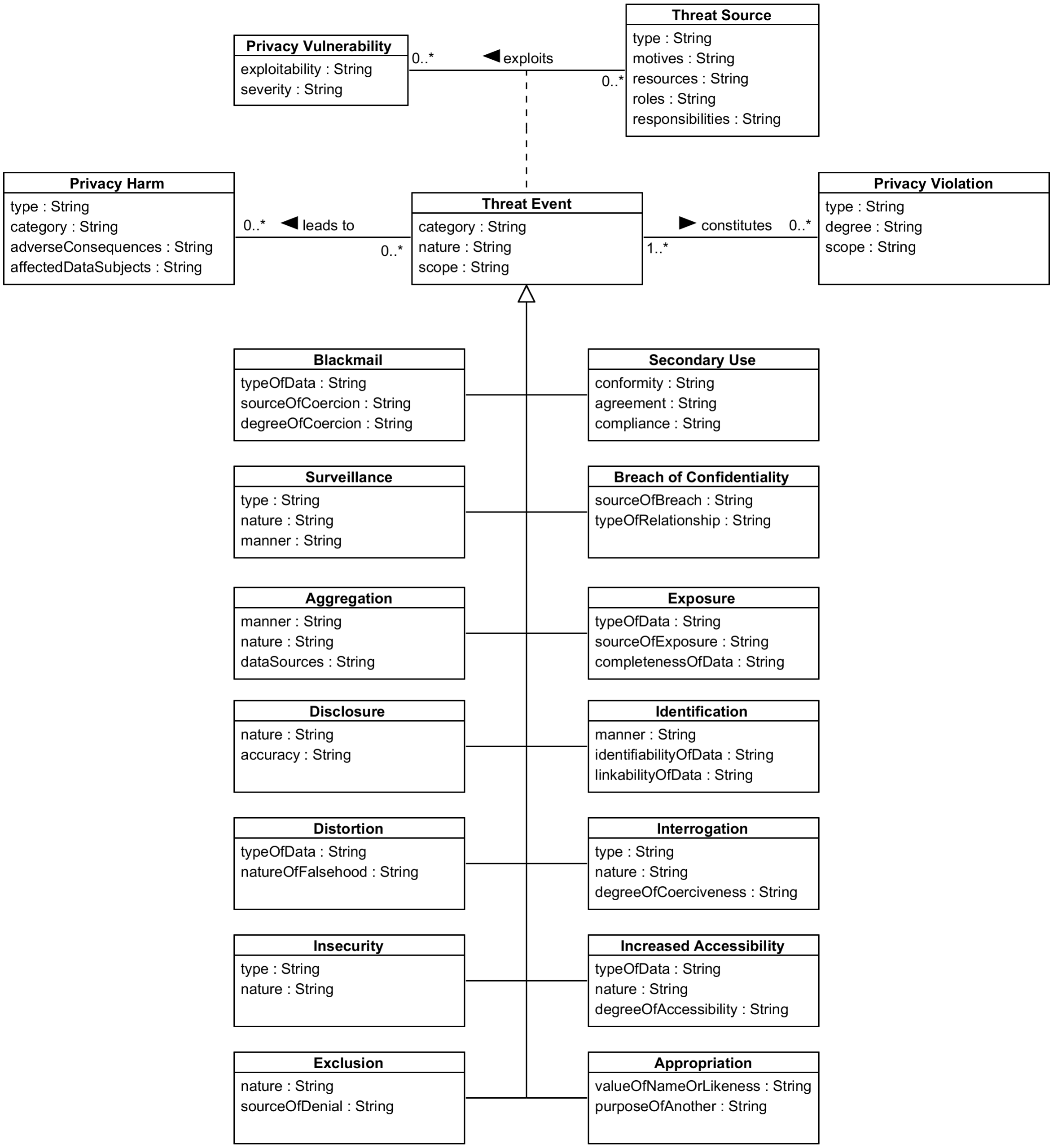

Fig. 1. The conceptual relationship among the key risk factors.

in eight stages: *Initiation, Collection, Retention, Access, Review, Usage, Disclosure* and *Destruction*. In each stage, data-processing activities and those required by legal frameworks and standards are concretely represented in *StageActivity, StageEvent* and *StageAction*. In addition, involved actors and the way in which they participate in processing activities are represented in *LifecycleRole* and *LifecycleActor*.

In order to describe the context in a widely-used modelling notation, we use the UML profile for the APDL model proposed in [24] to represent personal data, associated processes and involved actor in terms of the Unified Modeling Language (UML) [25].

B. Vulnerability Analysis

We assume that identifying and analysing vulnerabilities of the *supporting assets* is part of security risk analysis to ensure availability, integrity or confidentiality of the primary assets. We focus only on vulnerabilities of the *primary assets* to protect the privacy of data subjects and ensure the contextual integrity.

The first step in vulnerability analysis is to define a baseline model of processing that describes the targets of evaluation. Such a model is to understand reasonable expectations of privacy in each particular context. As such, we adopt the concept of contextual integrity [22], which was developed

TABLE I
DEPENDENCIES AMONG THE KEY FACTORS' ATTRIBUTES

Risk factors	Influenced attributes	Influencing attributes
Privacy vulnerability	Exploitability	Attributes of context-relevant processing norms (personal data, data-processing activities, involved actors, processing principles)
	Severity	Attributes of context-relevant processing norms (personal data, data-processing activities, involved actors, processing principles)
Threat source	Type	—
	Motives	Threat source (type) and data value derived from the attributes of personal data (type, category, sensitivity and linkability)
	Resources	Resources include skills, background knowledge, privileges, technical and financial resources, which depend on Threat source (type, roles and responsibilities)
	Roles	Threat source (type)
	Responsibilities	Threat source (role)
Threat event	Category	—
	Nature	Attributes of specialised threat events
	Scope	—
Privacy violation	Type	—
	Degree	Threat event (nature)
	Scope	Threat event (scope)
Privacy harm	Type	—
	Category	—
	Adverse consequences	Threat event (nature)
	Affected data subjects	Threat event (scope)

from social and philosophical theories to bring the social layer into view by identifying four main elements: contexts, attributes, actors and transmission principles. These elements constitute *context-relative informational norms*, which govern the flow of information in a particular context to ensure its appropriateness. From a technical perspective, these norms can be adapted by including processing operations as an element to consider both the flow of personal data and the processing of this data. In so doing, contextual integrity is about the appropriate flow and processing of personal data. Accordingly, we define context-relative processing norms as follows.

> In a context, the *processing* of a certain type (attributes) *of personal data* about a *data subject* (acting in a particular capacity/role) *by an actor (acting in a particular capacity/role)* is governed by a particular *processing* principle.

In order to comprehensively identify and analyse all possible vulnerabilities of the primary assets, a baseline model, which describes personal data, associated processes and involved actors, needs to be represented in a way that is amenable to analysis. As such, the baseline model of processing can be described in terms of context-relative processing norms. We adopt the APDL model as a source to capture and represent personal data, associated processing operations, involved actors and their assigned roles in each stage of the lifecycle. In addition, processing principles — which can be derived from legal frameworks, standards or domain-specific constraints — are represented as constraints for each data-processing activity in each stage of the data lifecycle. Similarly, we use the UML profile for the APDL model to describe the context-relative processing norms in a widely-used modelling notation.

Once the context and context-relative processing norms are established, vulnerabilities can be derived from how these norms would be breached or disrupted to violate contextual integrity. Crucially, each element of the norms (attributes, data-processing activities, actors and processing principles) need to be considered separately. Improper data model and a lack of data minimisation are examples of weaknesses for the element of *attributes* that may be exploited by a threat source which leads to the identification of a data subject as a threat event. For each vulnerability, its exploitability and severity need to be identified and estimated in relation to its attributes in Section IV-A2.

C. *Threat Analysis*

As mentioned in Section IV-A, a threat is an abstract concept that can be decomposed into two concrete concepts: threat sources and threat events.

1) Threat Sources: In order to identify all possible threat sources, it is necessary to establish the context in which personal data is collected and processed (as explained in Section V-A). The context helps support engineers in understanding the scope of analysis, multiple stakeholders, and the nature and sensitivity of the processed data. Once the context is established, a list of actors involved in the processing of personal data can be identified, along with assigned roles and responsibilities. In particular, the Initiation stage can be used to concretely identify the types of personal data to be collected and processed, and to abstractly identify involved actors and their roles and responsibilities. In order to identify involved actors at a detailed level of abstraction, we use the basic types of lifecycle roles (data modeller, data subjects, data controllers, data processors and third parties) in each stage of the lifecycle as a source of such details. A lifecycle role is a set of logically related activities that are expected to be conducted together and assigned to different actors as responsibilities according to their capabilities. In addition, a list of entities with interests or concerns in the value of these types of personal data can be identified. Data Protection Authorities, law enforcement bodies and other governmental agencies are examples of these entities. All such entities are potential threat sources. For each threat source, its type, motivation, resources, role and responsibilities need to be identified in relation to its attributes in Section IV-A.

2) Threat Events: Once the context is established, vulnerabilities and threat sources are identified, a list of events with the potential to adversely impact the privacy of data subjects can be identified. Most importantly, the identification of potential threat events need to be conducted in a systematic manner. As such, we adopt the taxonomy of privacy [21] as a means for characterising adverse privacy events. The taxonomy helps facilitate the identification of these events in a comprehensive and concrete manner. It classifies the most common adverse events into four basic groups: information collection, information processing, information dissemination and invasions. Adverse events are arranged around a model that begins with the data subject, from which various entities collect personal data. Data holders process — i.e. store, combine, manipulate, search and use — the collected data. They may also disseminate or release the processed data to other entities. Most importantly, the progression from collection through processing to dissemination is the personal data moving further away from the control of the data subject. In the last group of adverse events, i.e. invasions, the progression is toward the data subject and does not necessarily involve personal data [21].

We acknowledge that the taxonomy was developed to serve as a framework for the future development of the field of privacy law. This means that it covers all aspects and dimensions of privacy. In our approach, however, we focus only on those technical events that have implications on data privacy. From a technical perspective, these adverse events need to be arranged around a widely used model in the field of systems engineering or software engineering for describing the processing of data. The taxonomy classifies the most common adverse events into four basic groups that to a certain extent are arranged around a well-know processing model: the input-process-output (IPO) model. The first three groups — information collection, information processing and information dissemination — represent the input, process and output stages of the model respectively. In addition, the fourth group — invasions — is not related to that model as invasions are not only caused by technology and invasive adverse events do not always involve personal data, rather they directly affect data subjects. As such, we consider only some aspects of these events that involve personal data throughout the collection and disclosure stages of the lifecycle.

We use the IPO model as a starting point towards describing these events at a detailed level of abstraction. As such, we adopt the APDL as a model around which we arrange these events. Figure 3 illustrates the conceptual relationship between the taxonomy and the APDL model by mapping the basic groups of adverse events onto the stages of the data lifecycle.

Each type of an adverse threat event can be characterised by a set of attributes according to the nature of processing operations in each stage of the lifecycle. Threat events are data-driven events: they are more related to primary assets than supporting assets. Thus, threat events are identified in relation to the identified vulnerabilities.

- *Collection.* In the collection stage, adverse events are related to the manner in which personal data is collected: available choices and collection methods.
  - *Surveillance.* It involves collecting or recording a large amount of personal data about the data subject's activities. Each type of a threat event can be characterised by a set of attributes: manner (continuous or discrete monitoring), type (covert or overt — i.e. passive or active) and nature (extensive or limited).
  - *Interrogation.* It involves coercively collecting personal data by asking or probing unwarranted questions. Each type of a threat event can be characterised by a set of attributes: degree of coerciveness, type (direct or indirect) and nature (excessive or limited).
- *Retention.* In the retention stage, adverse events are related to the manner in which the collected personal data is structured, organised, stored and retained.
  - *Aggregation.* It involves structuring, organising, storing or retaining integrated items of personal data about a data subject. Each type of a threat event can be characterised by a set of attributes: manner (anticipated or unanticipated), nature (excessive or limited), and data sources (internal or external).
  - *Identification.* It involves structuring, organising, storing or retaining different items of personal data in a manner through which personal data can be linked to particular data subjects logically or physically. Each type of a threat event can be characterised by a set of attributes: manner (anticipated or unanticipated), identifiability of data (identified, pseudonymous or anonymous) and linkability of data to personal identifiers (linked, linkable with reasonable effort, not linkable with reasonable effort or unlinkable).
  - *Insecurity.* It involves improper data protection and handling. Each type of a threat event can be characterised by a set of attributes: nature (data handling or data protection) and type (design flaw, implementation flaw, retention time).
- *Access.* In the access stage, adverse events are related to the manner in which personal data is retrieved.
  - *Insecurity.* In this stage, handling includes retrieval mechanisms. Each type of a threat event can be characterised by a set of attributes: nature (data handling or data protection) and type (design flaw, implementation flaw).
- *Review.* In the review stage, adverse events are related

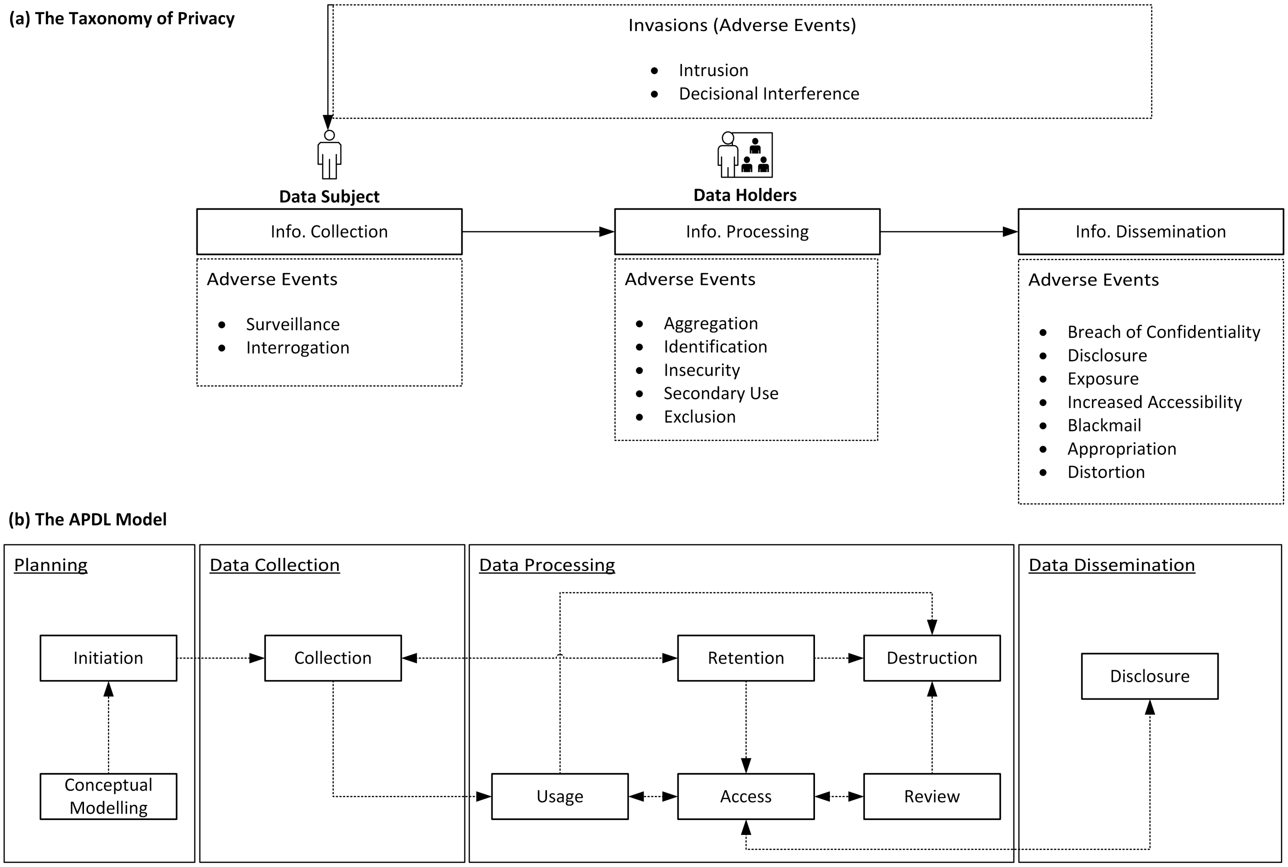

Fig. 3. The conceptual relationship between the taxonomy of privacy and the APDL model.

to the manner in which data subjects participate in the processing of personal by exercising their access rights to review or rectify their personal data and ensure that it is accurate, complete and up-to-date.

  - *Exclusion.* It involves the failure to provide data subjects with notice and access to their personal data. Each type of a threat event can be characterised by a set of attributes: nature (partial or complete) and source of denial (unjustified, necessary for processing, or required by law or regulation).

- *Usage.* In the usage stage, adverse events are related to the manner in which personal data is manipulated and used.
  - *Aggregation.* It involves altering, adapting, refining, aligning and combining or integrating different items of personal data about a data subject.
  - *Identification.* It involves altering, adapting, refining or aligning different items of personal data in manner through which personal data can be linked to particular data subjects.
  - *Insecurity.* It involves improper data protection and handling.
  - *Secondary Use.* It involves using the collected personal data for purposes unrelated to the purposes for which it was initially collected without the knowledge and consent of the data subject. Each type of a threat event can be characterised by a set of attributes: conformity (with the specified purposes), agreement (with the obtained consent) and compliance (with legal frameworks).
- *Destruction.* In the destruction stage, adverse events are related to the manner in which personal data is erased, destroyed, redacted or disposed.
  - *Insecurity.* It involves improper data protection and handling.
- *Disclosure.* In the disclosure stage, adverse events are related to the manner in which personal data is disseminated, made available or transmitted to third parties.
  - *Breach of confidentiality.* It involves the revelation of confidential personal data about a data subject by violating a trusted relationship. Each type of a threat event can be characterised by a set of attributes: source of breach (individual, group of individuals or institution) and type of relationship (trusted, semi-trusted or untrusted).
  - *Disclosure.* It involves the revelation of concealed and true personal data about a data subject to third parties. Each type of a threat event can be characterised by a set of attributes: nature (extensive or limited) and

accuracy (accurate or inaccurate).

- *Exposure.* It involves the revelation of concealed personal data that refers to physical or emotional attributes about a data subject. Each type of a threat event can be characterised by a set of attributes: type of data (physical or emotional attributes), source of exposure (individual, group of individuals or institution) and completeness of data (reflects the capability of data for judgement, whether it is complete and can be used for judging a data subject's personality or character).
- *Increased accessibility.* It involves making personal data that is already available to the public more easier to access. Each type of a threat event can be characterised by a set of attributes: type of data, nature (excessive or limited) and degree of accessibility.
- *Blackmail.* It involves coercing data subjects by threatening to reveal their concealed personal data for legal or illegal purposes. Each type of a threat event can be characterised by a set of attributes: type of data, degree of coercion and source of coercion.
- *Appropriation.* It involves the use of personal data that shapes a data subject's identity or personality for the purposes and goals of another. Each type of a threat event can be characterised by a set of attributes: value of the name or likeness (reputation, prestige, social or commercial standing) and purpose of another.
- *Distortion.* It involves exposing a data subject to the public inaccurately by revealing false and misleading personal data. Each type of a threat event can be characterised by a set of attributes: type of data and nature of falsehood (untrue, inaccurate or misleading data).

D. Privacy Harm Analysis

1) Privacy Violations: Once privacy vulnerabilities, threat sources and threat events are identified, privacy violations can be identified as illegitimate or unanticipated data-processing activities resulting from the occurrence of threat events without negative consequences on data subjects. In particular, for each possible exploitation, privacy violations are activities that can be conducted without adverse actions taken against data subjects, as well as without their knowledge. For each type of privacy violation, its degree and scope need to be identified in relation to its attributes in Section IV-A3.

2) Privacy Harms: Once privacy vulnerabilities, threat sources and threat events, privacy harms can be derived from these events as potential adverse consequences on the privacy of data subjects. Most importantly, legal and social factors that have impacts on the determination of privacy harms need to be considered. As such, we use the same categories of privacy harms of [1] that have been identified in previous attempts from a legal perspective [21] and [22]. In particular, privacy harms are classified into: physical; economic or financial harms; mental or psychological harms; harms to dignity or reputation; and societal or architectural harms [1]. We arrange these categories of harms around the APDL model according to its lifecycle stages, associated data-processing activities and their corresponding threat events, as illustrated in Table II. For each type of privacy harm, its type, adverse consequences and affected data subjects need to be identified in relation to its attributes in Section IV-A4.

VI. A CASE STUDY

In this section, we introduce the European Electronic Toll Service (EETS), the aim of which is to support interoperability between Electronic Toll Pricing (ETP) systems, and which we shall use to illustrate the applicability and usefulness of our approach in this particular context.

A. Overview

The European Electronic Toll Service (EETS) [26] aims to support interoperability between electronic road toll systems at a European level. The main purpose of collecting and processing EETS users' personal data is to electronically calculate and collect road-usage tolls. We have chosen this case study for the following reasons. First, it has been critically analysed with regards to privacy risks in the literature [5], [27]. Second, EETS is regulated by Directive 2004/52/EC of the European Parliament and of the Council of 29 April 2004 on the interoperability of electronic road toll systems in the Community [28] and the related Commission Decision 2009/750/EC of 6 October 2009 on the definition of the European Electronic Toll Service and its technical elements [29]. Third, the European Commission provides full details about EETS by publishing a guide as a reference manual for all parties concerned by the Directive and the Decision. The guide illustrates references and procedures to help the implementation of electronic road toll systems interoperability and EETS [26]. Finally, data-processing activities of the EETS have been already represented in [24].

EETS complements national electronic road toll systems to ensure their interoperability. It is intended to cover all domains and objects that are subject to toll, such as road networks, specific sections of roads (e.g. a bridge, a tunnel or a ferry connection), or specific areas offering services (e.g. a parking lot or access to a protected area in a city). It enables road users to easily pay road-usage tolls throughout the Member States with a single subscription contract with an EETS provider [26].

The main actors involved in the EETS are users, EETS providers and toll chargers. EETS providers are legal entities that grant access to EETS to road users [29]. Toll chargers are public or private organisations that are responsible for levying tolls for the circulation of vehicles in an EETS domain [29]. A user is an individual who subscribes to an EETS provider in order to get access to EETS, regardless of nationality, country of residence or the Member State in which the vehicle is registered [29]. By signing a contract, a user needs to provide a set of data — user and vehicle classification parameters — specified by a responsible toll charger, as well as to be informed about the processing of their personal data in relation to applicable law and regulations. Accordingly, the EETS provider provides the user with an On-Board Unit (OBU) to

TABLE II
PRIVACY HARMS IN RELATION TO THE STAGES OF THE APDL MODEL.

Threat event	Lifecycle stage	Category of harm	Description
Surveillance	Collection	Mental or psychological Societal or architectural	It creates feelings of anxiety and discomfort; it can lead to self-censorship and inhibition, which can adversely impact freedom of choice, creativity, and self-development. In addition, it can have a chilling effect on behaviour.
Interrogation	Collection	Mental or psychological Harms to dignity or reputation	It creates feelings of discomfort; it can lead to harm for conscience and human dignity.
Aggregation	Retention Usage	Mental or psychological Harms to dignity or reputation Societal or architectural	It can cause dignitary harms; it violates the expectations of data subjects by integrating data in potentially unanticipated manner.
Identification	Retention Access Usage	Mental or psychological Harms to dignity or reputation	It can cause dignitary harms, which stems from the way in which a data subject can be linked to a set of collected data. It inhibits data subjects' abilities to change and prevents their self-development and free expression.
Insecurity	Retention Access Usage Destruction	Economic or financial Mental or psychological Harms to dignity or reputation	It causes dignitary harms, which stems from judgement based on the defiled profiles by identity thieves, as well as from the lack of protection against data leakage.
Secondary use	Usage	Mental or psychological Harms to dignity or reputation	It can cause dignitary harms, which stems from using personal data without the knowledge and consent of data subjects.
Exclusion	Review	Mental or psychological	It can creates a sense of vulnerability and uncertainty in data subjects, which lead to lead to feelings of powerlessness.
Breach of confidentiality	Disclosure	Mental or psychological Harms to dignity or reputation	It violates the trust in the relationship, which leads to betraying data subjects regardless of the nature of the data revealed.
Disclosure	Disclosure	Physical Mental or psychological Harms to dignity or reputation	It can cause dignitary harms, which stems from disseminating embarrassed facts about data subjects. It can also threaten data subjects' security and make them vulnerable by using the disclosed data.
Exposure	Disclosure	Harms to dignity or reputation	It can cause dignitary harms, which stems from disseminating embarrassed and disgusting facts about data subjects.
Increased accessibility	Disclosure	Physical Economic or financial Mental or psychological Harms to dignity or reputation	It can increase the possibility of disclosure.
Blackmail	Disclosure	Societal or architectural	The harm is not in the actual disclosure of personal data, but in the control exercised by an individual who makes the threat over the data subject.
Appropriation	Disclosure	Harms to dignity or reputation	It can be harmful even it is not humiliating, degrading, as it shapes a person's identity. It also inhibits the freedom and self-development.
Distortion	Disclosure	Mental or psychological Harms to dignity or reputation Societal or architectural	It can result in embarrassment, humiliation, stigma, and reputational harm.

be installed on-board a vehicle to collect, store, and remotely receive and transmit time, distance and location data over time. This data, together with the user's and vehicle's parameters, are specified to declare the toll of circulating a vehicle in a specific toll domain [26]. EETS provision entails personal data processing, which must be carried out in compliance with the EU Directive 95/46/EC [30] and Directive 2002/58/EC [31].

B. An Illustration

Due to space limitations, we focus only on the main factors of privacy risks to illustrate the applicability of our approach rather than providing an exhaustive list of threat sources and events along with their corresponding violations and harms.

1) Context Establishment: All useful information that helps establish the context has been already captured by the APDL model in [24]. This includes the types of personal data; its sources; the purposes for, and the manner in which, these data types are processed; involved actors and their assigned roles and responsibilities; relevant legal frameworks and standards; and domain-specific constraints.

The establishment of the context in which personal data is collected and processed consists of three steps. The first step is to specify or model the types of personal data along with their attributes (captured by classes stereotyped by «PersonalData») and the main purpose for which this data is collected and processed (captured by a class stereotyped by «Purpose» along with its lawfulness, fairness and proportionality). With reference to the APDL model, the main purpose is to *electronically calculate and collect road-usage tolls* and the types of personal are:

- Identification and contact data — ***EETSUser:*** user ID, name, billing address (collected from the EETS's user whether the user is the driver, owner, lesser or fleet operator of the vehicle)
- Vehicle classification parameters — ***Vehicle:*** license plate, classification code (collected from the EETS user)
- Location data — ***LocationData:*** time, distance, place (collected from OBUs)

The second step is to specify or model process — both actual data-processing activities and privacy-related processes required by legal frameworks and standards — in each stage of the APDL model. These processes are abstractly captured from classes stereotyped by «Initiation», «Collection», «Retention», «Access», «Review», «Usage», «Disclosure» and «Destruction». With focus on location data, we illustrate a data-processing activity in the collection stage of the APDL model: it is abstractly captured from the ***CollectingUsageData*** class, which is stereotyped by «Collection». The stereotyped class also captures other important details: location data sources (OBUs), available choices (the user is entitled to subscribe to EETS with the EETS providers of their choice among other choices: the national or local manual, automatic or electronic toll services), collection method (OBUs using satellite positioning systems), consent type (implicit by signing a contract) and relevant GPS principles (Collection Limitation). In addition, processes are concretely captured from classes stereotyped by «StageActivity», «StageAction» and «StageEvent». Each stage activity contains a set of actions that represent its executable steps and a set of events that cause the execution of these actions. The data-processing activity is concretely captured from the ***CollectingLocationData*** class, which is stereotyped by «StageActivity». At this level of detail, it aims to collect road-usage data to be used for tolls declaration and calculation. The stereotyped class also captures other important details in terms of constraints: preconditions (the privacy notice needs to be communicated to EETS users at or before the collection time in a clear and concise manner; their implicit consent needs to be obtained at or before the collection time in an informed manner — i.e. the EETS user has already subscribed to the service; and the minimum necessary amount of location data needs to be modelled to fulfil the stated purpose) and post-conditions (the road-usage data has been successfully collected). This activity is decomposed into two classes: ***CollectLocationData*** and ***Collect***, which are stereotyped by «StageAction» and «StageEvent» respectively. CollectLocationData class captures the time of usage, the covered distance and the place on which the vehicle is circulating on a particular toll domain for tolls declaration and calculation. Collect class captures the occurrence of circulating a vehicle on a particular toll domain to collect location data.

The third step is to specify or model involved actors (captured by classes stereotyped by «LifecycleRole» and «LifecycleActor»). Each lifecycle stage includes a number of lifecycle roles, each of which is played by different actors according to their capabilities and responsibilities. With reference to the APDL model, ***CollectionAgent*** is a type of the data processor role that consists of logically related activities for collecting road usage data, and ***ServiceProvider*** is a type of involved actors who are capable of, and responsible for, performing the activities of the collection agent as a role to which are assigned. Responsibilities are captured from stage activities in which a lifecycle actor participates and to which a lifecycle role associates.

Establishing the context in which personal data is collected and processed requires specifying or modelling the *primary assets* along with *involved actors* and their roles and responsibilities. The APDL model has served as a preliminary acquisition step to capture all required data that support privacy risk analysis and compliance checking.

2) Vulnerability Analysis: In our approach the focus is on vulnerabilities of *primary assets* to protect the privacy of data subjects and ensure the contextual integrity. The first step of vulnerability analysis is to develop a baseline model of the processing of personal data. The baseline model captures all appropriate data-processing activities in all stages of the APDL model. In order to develop a baseline model, we need to establish a context-relative processing norm for each processing of a type of personal data. The main elements that constitute these norms are captured from stage activities in the established context. Due to space limitations, we identify only a context-relative processing norm for the *CollectingLocationData* activity illustrated in Section VI-B1.

> In the context of EETS, the collection of a certain type of personal data (location data: time, distance, place) about an EETS user (acting as a data subject) by an EETS provider (acting as a data processor on behalf of a toll charger) is governed by processing principles derived from applicable legal frameworks ([28], [29], [30], [31]) and standards (relevant GPS principles).

In this case, legal framework principles — for example, DIRECTIVE 95/46/EC — are as follows. Personal data must be

- processed fairly and lawfully,
- collected for specified, explicit and legitimate purposes,
- adequate, relevant and not excessive, and
- accurate and up to date.

In addition, the relevant GPS principle is *Collection Limitation*. Most importantly, principles of legal frameworks and standards are modelled as constraints in the APDL model — i.e. pre and post-conditions for each stage activity.

Once all context-relevant processing norms are defined in relation to the APDL model, a complete baseline model can be developed to serve as the basis for deriving privacy vulnerabilities. The second step of vulnerability analysis is to derive all possible vulnerabilities of the primary assets from the identified context-relevant processing norms. They can be derived by examining all the main elements that constitute each processing norm — i.e. any possible breach of a processing norm can be derived as a vulnerability. With reference to the above processing norm, a possible vulnerability with regards to *attributes*, as an element, is 'an improper data model' (PV.1) that directly or indirectly links location data to users' IDs. Another possible vulnerability with regards to *processing principles*, as an element, is 'a lack of data minimisation' (PV.2) that facilitates inadequate, irrelevant and excessive collection of location data in an interval basis, which is not necessary for the main purpose. In addition, 'improper purpose specification' (PV.3), 'an improper prefer-

TABLE III
THE MOST IMPORTANT VULNERABILITIES IN THE CONTEXT OF EETS.

Code	Privacy vulnerability
PV.1	An improper data model
PV.2	A lack of data minimisation
PV.3	An improper purpose specification
PV.4	An improper preference specification
PV.5	A weak anonymisation technique
PV.6	A lack of logs and audit trails

ence specification' (PV.4) and 'a lack of logs and audit trails' (PV.6) are other types of vulnerabilities relating to processing principles. Furthermore, 'a weak anonymisation technique' (PV.5) facilitates the re-identification of particular EETS users and the inference of sensitive data. Table III shows the most important privacy vulnerabilities in the context of EETS.

3) Threat Analysis: Due to space limitations, we do not consider an exhaustive list of all types of threat sources and threat events; rather, we identify only the most important types of those to illustrate the applicability and usefulness of our approach.

a) Threat Sources: In reference to the established context, EETS providers (TS.1) are involved in the processing of EETSUser and LocationData by playing the role of data processors who grant access to EETS to EETS users. They may act accidentally or deliberately as threat sources while they process personal data lawfully to calculate and communicate personalised fees (road-usage tolls) for each EETS user by the end of the tax period — or unlawfully for further processing with the motivation of profiling EETS users, discriminatory social sorting or providing better services. The utility of 'location data' and 'identification and contact data' in this context makes such data highly valuable to EETS providers. The value of this data stimulates the motives of EETS providers to exploit vulnerabilities of the primary assets. In particular, it has a market value when it is exploited by EETS providers for administrative and commercial purposes — for example, it gives an EETS provider a competitive advantage with respect to their competitors. According to the attributes of a threat source, EETS providers are insiders and institutions. EETS providers have technical skills and detailed background knowledge about conceptual, logical and physical data models, as well as about the processing operations. It also implies that they have legitimate privileges to collect and process location-related data according to their roles and responsibilities. Based on these, they have access rights to both the 'fine-grained location data' and 'identification and contact data'. In addition, they have reasonable resources — both technical and financial — to get benefit from the values of the collected data by creating comprehensive and identifiable profiles.

Similarly, toll charger (TS.2) are involved in the processing of EETSUser and LocationData by playing the role of data controllers who are responsible for levying tolls for the circulation of vehicles in an EETS domain. They may act accidentally or deliberately as threat sources while they process EETSUser and LocationData lawfully to have access to evidence proving that a vehicle was at a specific location at a particular time for exception handling or enforcement support — e.g. a photograph taken by a road-side equipment or a toll gate, or unlawfully for further processing with the motivation of profiling EETS users, discriminatory social sorting, managing blacklists, or providing complaint resolutions. The value of EETSUser and LocationData stimulates the motives of toll chargers to exploit vulnerabilities of the primary assets. In particular, it has a market value when it is exploited by toll chargers for administrative and commercial purposes. According to the attributes of a threat source, toll chargers are insiders and institutions. Toll chargers have technical skills and detailed background knowledge about conceptual, logical and physical data models, as well as about the processing operations. They also have legitimate privileges to access and process location-related data according to their roles and responsibilities. Based on these, they have access rights to both the 'fine-grained location data' and 'identification and contact data'. In addition, they have reasonable resources — both technical and financial — to get benefit from the values of the collected data by creating comprehensive and identifiable profiles.

According to the nature and utility of LocationData and EETSUser specified in the established context, it is obvious that there is a number of external entities who are not directly involved in the processing of this data; rather, they may act as third parties with interests or concerns in the value of these types of personal data with various motives and resources.

- Departments, agencies and public bodies. Fine-grained location data may be of interest of state agencies for several motives.
  - Department for Transport (TS.3): Location data can be used as a source for collecting traffic statistics to improve road mobility by applying congestion charges.
  - Intelligence and security services (TS.4): Location data can be used to facilitate law enforcement investigations by discovering whether individuals are where they claim to have been at any point in time. It can also be used to identify and put individuals under surveillance based on their associations with others or the locations frequented.
- Employment agencies (TS.5): Location data can be used for background checking. Those agencies may make excessive inference with the aim of, for example, deriving health conditions from driving patterns. The agencies may use the derived data for filtering job candidates based on these conditions.
- Health insurance providers (TS.6): Location data can be used to make excessive inference with the aim of deriving health conditions. Those conditions are considered as one of the main factors for calculating health insurance premiums.
- Car insurance providers (TS.7): Location data can be used to discover whether a driver is where they claim to have

TABLE IV
THE MOST IMPORTANT THREAT SOURCES IN THE CONTEXT OF EETS.

Code	Threat source
TS.1	EETS providers
TS.2	Toll chargers
TS.3	Department for Transport
TS.4	Intelligence and security services
TS.5	Employment agencies
TS.6	Health insurance providers
TS.7	Car insurance providers
TS.8	Advertising companies

been at any point in time. In addition, it can be used to make excessive inference with the aim of deriving driving patterns, which are one of the main factors for calculating car insurance premiums.

- Advertising companies (TS.8): Location data can be used to make excessive inference with the aim of, for example, deriving health conditions or religious beliefs. These companies may use the derived data for sending targeted advertising or unsolicited emails.

Table IV shows the most important threat sources in the context of EETS.

b) Threat Events: In a straightforward implementation of the EETS architecture, the calculation of road-usage tolls is performed remotely at EETS providers' back-office systems. The OBU collects, stores, and remotely receives and transmits time, distance and place over time to the EETS provider's back-office systems. These systems are in charge of processing location data to calculate personalised road-usage tolls and communicate the final premium to EETS user at the end of the tax period. As mentioned, a threat event occurs as a result of a successful exploitation of one or more vulnerabilities by one or more threat sources. With reference to the identified vulnerabilities and threat sources, we identify the most significant threat events with the potential to adversely impact the privacy of EETS users that may happen at specific points in time. The identification of these events needs to be conducted in a systematic manner — i.e. according to the stages of the data lifecycle.

In the collection stage, threat events that may lead to privacy violations or harms are related to the manner in which personal data is collected. By exploiting PV.2, TS.1 may use OBUs to excessively collect irrelevant location data (TE.1) in fine-grained manner about EETS users. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'surveillance'. It is characterised as continuous, overt and extensive: continuous via the collection of location data over time; overt via informing the EETS user about the manner in which location data will collected when signing the contract; and extensive via the excessive collection of location data in a fine-grained manner throughout national and international toll domains. Surveillance outside toll domains implicates reasonable expectations of privacy as it may reveal hidden details that would not ordinarily be observed by others.

In the retention stage, threat events that may lead to privacy violations or harms are related to the manner in which personal data is structured, organised, stored and retained. By exploiting PV.1, TS.1 may store driving profiles for EETS users (TE.2) by integrating multiple items of personal data from various sources. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'aggregation'. It is characterised as unanticipated and excessive: unanticipated via the failure of informing EETS users about potential integration of additional data items; and excessive via the integration of fine-grained location data collected over time into 'identification and contact data' (internal sources), and 'vehicle classification parameters' (external sources). Location data alone does not reveal much more knowledge, but combining or integrating different items of personal data can reveal new facts about EETS users that they did not expect would be known when the original data was collected. EETS users' driving profiles can be created as a result of aggregation and may be used for judgement or evaluation of EETS users' financial reputation as it reveals credit history of previous tolls' payments.

In addition, by exploiting PV.1, TS.1 may store identifiable driving profiles for EETS users (TE.3) by linking driving profiles to particular EETS users. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'identification'. It is characterised as identified and linked: identified via structuring and storing identification and contact data — i.e. user ID, name and billing address — and linked via storing contact data with location data or using unique identifiers across databases — i.e. user ID, contract serial number and account number.

In the usage stage, threat events that may lead to privacy violations or harms are related to the manner in which personal data is classified, analysed, manipulated and used. By exploiting PV.1, TS.1 may create identifiable driving profiles for EETS users (TE.4) by linking driving profiles to particular EETS users. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'identification'. It is characterised as identified via integrating identifiable location-related data, and linked via integrating 'identification and contact' data with 'location data' or using unique identifiers across databases.

In addition, by exploiting PV.1 and PV.2, TS.1 may excessively infer sensitive information (TE.5) by analysing the aggregated data — EETS users' profiles — in a particular data analysis technique, for example, data mining to discover useful information, such as driving patterns that may reveal health conditions among others. With reference to the adapted taxonomy of adverse privacy events, the intensive inference event is a type of 'identification'. It is characterised as identified and linked: identified via integrating identification and contact data — i.e. user ID, name and billing address — and linked via integrating contact data with location data or using unique identifiers across databases — i.e. user ID, contract serial number and account number.

By exploiting PV.3, PV.4 and PV.6, TS.1 may use EETS users' profiles for further processing (TE.6). This includes commercial or malicious purposes which are not related to the purposes for which location data was initially collected and for which EETS users have provided implicit or explicit consent. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'secondary use'. It is not: in conformity with the specified purposes; in agreement with the obtained consent; or in compliance with applicable legal frameworks.

In the disclosure stage, threat events that may lead to privacy violations or harms are related to the manner in which personal data is disseminated, made available or transmitted to third parties for external use. By exploiting PV.4 and PV.6, TS.1 may share driving patterns with third parties (TE.7) that are with interests or concerns in the value of these types of personal data. With reference to the adapted taxonomy of adverse privacy events, this threat event is a type of 'disclosure'. It is characterised as an extensive disclosure of accurate data.

The utility of location data can be of interest of TS.4, TS.5, TS.6, TS.7 or TS.8 for several purposes by exploiting PV.5. TS.4 may excessively make inference (TE.8) with the aim of re-identifying data subjects to facilitate law enforcement investigations. In addition, TS.4 may re-identify data subjects to put them under close surveillance based on their associations with others or the locations frequented. TS.5 may excessively make inference (TE.8) with the aim of re-identifying data subjects for background checking to filter job candidates based on the derived health conditions. TS.6 may excessively make inference (TE.8) with the aim of re-identifying data subjects to use the derived health conditions as a criterion for calculating health insurance premium. TS.7 may excessively make inference (TE.8) with the aim of re-identifying data subjects to use the derived vehicle use and health conditions as criteria for calculating car insurance premium. TS.8 may excessively make inference (TE.8) with the aim of re-identifying data subjects to use the derived religious beliefs or health conditions as references for sending targeted advertising or unsolicited emails. With reference to the adapted taxonomy of adverse privacy events, this event is a type of 'identification'. It is characterised as an unanticipated event that uses anonymised data and this data is linkable with reasonable effort.

Table V summarises the above threat events along with the corresponding threat sources and privacy vulnerabilities.

4) Harm Analysis: In this section, we illustrate the most significant privacy violations and harms that may result from the occurrence of the threat events illustrated in Section VI-B3.

a) Privacy Violations: In the collection stage, 'passive collection of location data outside toll domains' is a privacy violation that may result from the occurrence of the threat event 'excessive collection of location data', which results from the successful exploitation of 'a lack of data minimisation' by EETS providers. Its degree is excessive as it collects fine-grained location data outside toll domains, whether they are national and international. Its scope is individuals — i.e. those who are subscribed to EETS. This privacy violation is considered as an illegitimate and unanticipated data-processing activity without adverse consequences. In particular, fine-grained location data is collected in ways EETS users would not reasonably expect, as well as this data is collected passively without the knowledge and consent of EETS users. In addition, the collection of location data outside toll domains does not have legitimate grounds as they are irrelevant and inadequate for the purposes for which location data is collected. Most importantly, this privacy violation is assumed to be without adverse actions taken against EETS users.

In the retention stage, 'unjustified retention' of location data is an example of a privacy violation that may result from the occurrence of the threat event 'exceeded retention period', which results from the successful exploitation of 'inappropriate retention schedule' and 'a lack of logs and audit trails' by EETS providers. Its degree is excessive as it retains the fine-grained location data for longer than necessary that exceed the specified retention time without operational or legal justifications. Its scope is individuals — i.e. those who are subscribed to EETS. This privacy violation is considered as illegitimate and unanticipated data-processing activity without adverse consequences. In particular, location data is retained in ways EETS users would not reasonably expect, as well as it is retained beyond the specified retention time and without the knowledge and consent of EETS users. In addition, the retention of location data does not have operational and legitimate grounds as it is no longer necessary to fulfil the specified purposes. Most importantly, this privacy violation is assumed to be without adverse actions taken against EETS users.

b) Privacy Harms: Privacy harm analysis is the most important step of any privacy risk-analysis approach. Privacy harms are derived from the undesirable consequences of threat events as potential adverse actions taken against data subjects. In this report, we consider only the objective category of privacy harms as the subjective category is mainly about the perception of unwanted observation.

For each stage of the data lifecycle, the potential undesirable consequences of each threat event need to be identified. Then, these consequences need to be analysed to determine whether they can partially contribute to, or completely lead to a negative action that uses personal data against the data subject in an unanticipated or coerced manner. Most broadly, a privacy harm may result from a series of adverse consequences of multiple threat events.

In the collection stage, the main undesirable consequence of TE.1 is gathering a large amount of fine-grained location data that has been collected over time as comprehensive driving records (UC.1), which may include complete driving history or driving history for a specific period for EETS users.

In the retention stage, the main undesirable consequence of TE.3 is storing identifiable driving records for EETS users (UC.2), from which driving profiles can be derived.

In the usage stage, the main undesirable consequence of TE.4 is creating identifiable driving profiles for EETS users (UC.3), from which driving patterns can be derived. In ad-

TABLE V
A SET OF REASONABLE EXPLOITATIONS.

Code	Threat event	Threat source	Privacy vulnerability
TE.1	Excessive data collection	TS.1	PV.2
TE.2	Aggregated data retention	TS.1	PV.1
TE.3	Identifiable data retention	TS.1	PV.1
TE.4	Identifiable data integration	TS.1	PV.1
TE.5	Unjustified data re-identification	TS.1	PV.1 and PV.2
TE.6	Unauthorised secondary use	TS.1	PV.3, PV.4 and PV.6
TE.7	Unauthorised data disclosure	TS.1	PV.4 and PV.6
TE.8	Excessive data collection	TS.4, TS.5, TS.6, TS.7 and TS.8	PV.5

TABLE VI
THE MOST IMPORTANT UNDESIRABLE CONSEQUENCES OF THREAT EVENTS.

Code	Undesirable consequence	Relevant threat event
UC.1	Comprehensive driving records	TE.1
UC.2	Identifiable driving records	TE.3
UC.3	Identifiable driving profiles	TE.4
UC.4	Identifiable driving patterns	TE.5
UC.5	Anonymised driving patterns	TE.7
UC.6	Sensitive facts about EETS users	TE.8

dition, the main undesirable consequence of TE.5 is deriving identifiable driving patterns (UC.4) that may be analysed to infer new sensitive facts about EETS users, such as occupation, geographical residence, religious beliefs, health conditions or political affiliation.

In the disclosure stage, the main undesirable consequence of TE.7 is revealing anonymised driving patterns for subscribed EETS users beyond expected boundaries (UC.5) that may be analysed to derive sensitive information that can inhibit certain rational judgements. In addition, the main undesirable consequence of TE.8 is deriving sensitive information about EETS users from their driving patterns (UC.6), based on which adverse actions against the data subjects can be taken by the relevant threat sources.

Table VI shows the most important undesirable consequences of TE.1, TE.3, TE.4, TE.5, TE.7 and TE.8 in the context of EETS. By analysing UC.4, UC.5 and UC.6, together with the relevant privacy vulnerabilities and threat sources, we can derive six privacy harms as follows.

- *Increased car insurance premium.* It occurs as EETS providers can make excessive inference to derive EETS users' driving patterns and share anonymised patterns with car insurance providers. Insurance providers may make inference to re-identify current and potential customers with the aim of calculating car insurance premium based on the types of vehicle use and health conditions, which are derived from their driving patterns.
- *Increased health insurance premium.* It occurs as EETS providers can make inference to derive EETS users' driving patterns and share anonymised patterns with health insurance providers. Insurance providers may make inference to re-identify current and potential customers with the aim of calculating health insurance premium based on health conditions, which are derived from their driving patterns.
- *Denial of a job.* It is a type of employment discrimination that may occur as EETS providers can make inference to derive EETS users' driving patterns and share anonymised patterns with employment agencies. Those agencies may make inference to re-identify job applicants with the aim of filtering those job candidates according to their health conditions or religious beliefs, which are derived from their driving patterns.
- *Being under close surveillance.* It occurs as EETS providers can aggregate fine-grained location data as driving records which may include complete driving history or driving history for a specific period for EETS users and link these records to 'identification and contact data' with the aim of creating identifiable driving profiles. In addition, a special type of intrusion may be performed by those who may be interested in identifying and putting under surveillance a number of EETS users based on their associations with others or the locations frequented that are derived from their driving patterns.
- *Receipt of targeted advertising.* It occurs as EETS providers can make inference to derive EETS users' driving patterns and share anonymised patterns with advertising companies without implicit or explicit consent. Those companies may make excessive inference to re-identify data subjects and send targeted advertising emails that implicitly make reference to their religion beliefs or health conditions.
- *Receipt of unsolicited mails.* It occurs as EETS providers can make inference to derive EETS users' driving patterns and share anonymised patterns with advertising companies without implicit or explicit consent. Those companies may send unsolicited emails for commercial purposes based on health conditions, religious beliefs, etc., which are derived from their driving patterns.

Table VII shows the most significant privacy harms, along with associated threat sources, privacy vulnerabilities, threat events and undesirable consequences of these events.

VII. CONCLUSIONS

We have presented a privacy risk model that considers organisational, legal, social and technical aspects of privacy.

TABLE VII
THE MOST SIGNIFICANT PRIVACY HARMS.

Code	Privacy harm	Threat source	Privacy vulnerability	Threat event	Undesirable consequences
PH.1	Increased car insurance premium	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.7	PV.5	TE.8	UC.6
PH.2	Increased health insurance premium	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.6	PV.5	TE.8	UC.6
PH.3	Denial of a job	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.5	PV.5	TE.8	UC.6
PH.4	Being under close surveillance	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.4	PV.5	TE.8	UC.6
PH.5	Receipt of targeted advertising	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.8	PV.5	TE.8	UC.6
PH.6	Receipt of unsolicited emails	TS.1	PV.1 and PV.2	TE.5	UC.4
		TS.1	PV.4 and PV.6	TE.7	UC.5
		TS.8	PV.5	TE.8	UC.6

In particular, we refer to fundamentals from the legal privacy literature to refine the meaning and properties of the key terms and risk factors, as well as the conceptual relationships between these factors. These fundamentals help support the distinction between privacy harms and violations by providing specific boundaries and properties, which are necessary for identifying and analysing privacy harms. In addition, the fundamentals bring legal and social layers into consideration by defining context-relative informational norms, from which context-relevant processing norms can be defined. These norms can be used to define a baseline model of processing, from which privacy vulnerabilities can be identified. The fundamentals also facilitate the identification of threat events in a systematic manner by providing a taxonomy of harmful activities and their corresponding harms. They also support the taxonomy by providing two main principles: (1) the limiting principle to help protect against reduction of the concept of privacy, and (2) the rule of recognition to support the identification of novel privacy harms as they emerge.

In addition, we have presented an analysis approach that describes how combinations of risk factors are identified and analysed at an appropriate level of detail. To improve the effectiveness of the analysis, our analysis approach complements a threat-oriented analysis approach by an asset/impact-oriented analysis approach. In particular, it starts with the identification of the primary assets. For each type of these assets, it identifies all possible vulnerabilities and the most likely threat sources who may exploit them. Based on these, it identifies the possible threat events through which the threat sources may exploit these vulnerabilities according to the types of the primary assets. In the context of threats, it identifies privacy violations and harms based on the undesirable consequences and potential adverse actions of these threat events. Furthermore, we have used tables to present traceable links between the key risk factors for each possible threat scenario. However, it is analytically useful to adopt an appropriate analysis technique that provides an effective way for considering the relationships among the key risk factors from which a reasonable set of threat scenarios can be generated.

We will build upon this work in a number of ways. First, we will use additional case studies to further validate the approach and highlight its usefulness and practical impact in various domains. Second, we intend to adopt an appropriate analysis technique that considers the many-to-many relationships among the key risk factors. Such an analysis technique will help identify, analyse and assess a reasonable set of threat scenarios, from which the most important vulnerabilities that need to be addressed can be identified. We also plan to identify an assessment approach to underpin a systematic risk-assessment methodology that can complement PIA processes. Such a methodology can be used as a means for managing the identified privacy risks in a structured manner by determining and reasoning about the most appropriate privacy controls for a particular context.

REFERENCES

[1] S. Joyee De and D. Le Métayer, "Priam: A privacy risk analysis methodology," in *11th International Workshop on Data Privacy Management and Security Assurance*. Springer, 2016, pp. 221–229.

[2] D. Wright, "The State of the Art in Privacy Impact Assessment," *Computer Law & Security Review*, vol. 28, no. 1, pp. 54–61, 2012.

[3] The European Union: Official Journal of the European Union, "General Data Protection Regulation," http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN, 2016.

[4] P. Dourish and K. A., "Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena." *Human–Computer Interaction*, vol. 21, no. 3, pp. 319–342, 2006.

[5] S. Gürses, C. Troncoso, and C. Diaz, "Engineering Privacy by Design," *Computers, Privacy & Data Protection*, vol. 14, no. 3, p. 25, 2011.

[6] A. Cavoukian, S. Shapiro, and R. J. Cronk, “Privacy Engineering: Proactively Embedding Privacy, by Design,” https://www.privacybydesign.ca/content/uploads/2014/01/pbd-priv-engineering.pdf, 2014.

[7] A. Cavoukian, “Privacy by Design,” https://www.privacybydesign.ca/content/uploads/2009/01/privacybydesign.pdf, 2009.

[8] S. Spiekermann, “The Challenges of Privacy by Design,” *Communications of the ACM*, vol. 55, no. 7, pp. 38–40, 2012.

[9] Cavoukian, A., “Privacy by Design: The 7 Foundational Principles Implementation and Mapping of Fair Information Practices,” https://www.ipc.on.ca/english/Resources/Discussion-Papers/Discussion-Papers-Summary/?id=953, 2010.

[10] A. Cavoukian, M. Monica, A. Fariba, R. Dan, and K. Jeff, “Privacy Risk Management: Building privacy protection into a Risk Management Framework to ensure that privacy risks are managed, by default,” https://www.ipc.on.ca/images/Resources/pbd-priv-risk-mgmt.pdf, 2010.

[11] M. C. Oetzel and S. Spiekermann, “A systematic methodology for privacy impact assessments: a design science approach,” *European Journal of Information Systems*, vol. 23, no. 2, pp. 126–150, 2014.

[12] R. Clarke, “Privacy impact assessment: Its origins and development,” *Computer Law and Security Review: The International Journal of Technology and Practice*, vol. 25, no. 2, pp. 123–135, 2009.

[13] D. Wright, K. Wadhwa, P. De Hert, and D. Kloza, “A Privacy Impact Assessment Framework for data protection and privacy rights,” http://www.piafproject.eu/Deliverables.html, 2011.

[14] A. Fineberg and P. Jeselon, “A Foundational Framework for a Privacy by Design — Privacy Impact Assessment,” http://privacybydesign.ca/content/uploads/2011/11/PbD-PIA-Foundational-Framework.pdf, 2011.

[15] R. Clarke, “An evaluation of privacy impact assessment guidance documents,” *International Data Privacy Law*, vol. 1, no. 2, pp. 111–120, 2011.

[16] BSI (Bundesamt für Sicherheit in der Informationstechnik), “IT-Grundschutz-Kataloge,” https://www.bsi.bund.de/DE/Themen/ITGrundschutz/ITGrundschutzKataloge/itgrundschutzkataloge_node.html, 2011.

[17] ——, “Risk analysis on the basis of IT-Grundschutz, BSI Standard 100-3,” https://www.bsi.bund.de/EN/TheBSI/thebsi_node.html, 2008.

[18] Commission Nationale de l´Informatique et des Libertés (CNIL), “Methodology for Privacy Risk Management,” https://www.cnil.fr/sites/default/files/typo/document/CNIL-ManagingPrivacyRisks-Methodology.pdf, 2016.

[19] ISO, “ISO 31000 - Risk management – Principles and guidelines,” http://www.iso.org/iso/home/standards/iso31000.htm, 2009.

[20] M. R. Calo, “The Boundaries of Privacy Harm,” *Indiana Law Journal*, vol. 86, pp. 1131–1162, 2011.

[21] D. J. Solove, “A Taxonomy of Privacy,” *University of Pennsylvania Law Review*, vol. 154, no. 3, pp. 477–564, 2006.

[22] H. F. Nissenbaum, *Privacy in context: Technology, Policy, and the Integrity of Social Life*. Stanford University Press, 2009.

[23] M. Alshammari and A. C. Simpson, “Personal Data Management: An Abstract Personal Data Lifecycle Model,” in *Proceedings of SPBP’17: Workshop on Security and Privacy-enhanced Business Process Management*. Springer, 2017.

[24] ——, “A UML Profile for Privacy-Aware Data Lifecycle Models,” in *Proceedings of the 1st International Workshop on SECurity and Privacy Requirements Engineering (SECPRE 2017)*. Springer, 2017.

[25] Object Management Group, “OMG Unified Modeling Language (OMG UML),” http://www.omg.org/spec/UML/, 2015.

[26] The European Commission, “The European Electronic Toll Service (EETS): 2011 Guide for the Application of the Directive on the Interoperability of Electronic Road Toll Systems,” http://ec.europa.eu/transport/themes/its/road/application_areas/electronic_pricing_and_payment_en, 2011.

[27] J. Balasch, A. Rial, C. Troncoso, B. Preneel, I. Verbauwhede, and C. Geuens, “PrETP: Privacy-Preserving Electronic Toll Pricing,” in *Proceedings of the 19th USENIX Security Symposium*, 2010, pp. 63–78.

[28] The European Union: Official Journal of the European Communities, “Directive 2004/52/EC of the European Parliament and of the Council of 29 April 2004 on the interoperability of electronic road toll systems in the Community,” http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32004L0052R(01)&from=EN, 2004.

[29] The European Union: Official Journal of the European Communities, “Commission Decision 2009/750/EC of 6 October 2009 on the definition of the European Electronic Toll Service and its technical elements,” http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32009D0750&from=EN, 2009.

[30] The European Union: Official Journal of the European Communities, “Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data,” http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=EN, 1995.

[31] The European Union: Official Journal of the European Communities, “Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector,” http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32002L0058&from=EN, 2002.