Navigating Privacy Patterns in the Era of Robotaxis

Ala'a Al-Momani
Ulm University
Ulm, Germany
alaa.al-momani@uni-ulm.de

David Balenson
USC Information Sciences Institute
Marina del Rey, CA, USA
balenson@isi.edu

Zoltán Ádám Mann
University of Amsterdam
Amsterdam, The Netherlands

Sebastian Pape
Continental Automotive Technologies GmbH
Frankfurt, Germany
sebastian.pape@continental.com

Jonathan Petit
Qualcomm Technologies Inc.
Boxborough, MA, USA
petit@qti.qualcomm.com

Christoph Bösch
Bosch Research
Renningen, Germany
christoph.boesch@de.bosch.com

Abstract—**Privacy engineering encompasses various methodologies and tools, including privacy strategies and privacy patterns, aimed at achieving systems that inherently respect privacy. Despite the collection of numerous privacy patterns, their practical application remains under-explored. This paper investigates the applicability of privacy patterns in the context of robotaxis, a use case in the broader Mobility-as-a-Service (MaaS) ecosystem. Using the LINDDUN framework for privacy threat elicitation, we analyze existing privacy patterns to address identified privacy threats. Our findings reveal challenges in applying these patterns due to inconsistencies and a lack of guidance, as well as a lack of suitable privacy patterns for addressing several privacy threats. To fill the gaps, we propose ideas for new privacy patterns.**

1. Introduction

Privacy engineering encompasses a set of methodologies and tools developed to achieve systems that inherently respect privacy. Notably, *privacy strategies* and *privacy patterns* [17] were introduced to identify suitable measures to enhance privacy and reduce privacy risks. While privacy strategies are high-level design guidelines intended to help meet data protection requirements, privacy patterns describe design solutions to common privacy problems and are considered "best practices" in privacy engineering. The community has collected a large number of privacy patterns[1]. However, only limited research addresses the useability of these patterns [8], [9].

The aim of this paper is to explore the applicability and usefulness of privacy patterns to improve privacy in future automotive systems. To this end, we consider the use case of robotaxis, in which a connected and automated vehicle (CAV) transports a rider from a pick-up point to a drop-off point. Robotaxis research has so far received limited attention to a systematic understanding of privacy threats and solutions to address these threats. We conduct a systematic privacy threat elicitation using LINDDUN [14] and then analyze to what extent existing privacy patterns help to find solutions to the identified threats.

The use case of robotaxis is particularly interesting as it consists of CAVs as part of a larger Mobility-as-a-Service (MaaS) ecosystem [22] involving various stakeholders, each with their own privacy requirements. A MaaS ecosystem enables users to obtain information, plan, book, and pay for a variety of mobility services via a single platform [22]. CAVs are a promising technology for the future of transportation, enabling driverless riding and improved road safety [24]. CAVs are equipped with a diverse set of sensors (e.g., forward-facing camera, in-vehicle camera, radar, lidar) to monitor, detect, understand, and react to events within the vehicle and on the road. CAVs generate, store, and send data to service providers to ensure adequate quality of service. However, the management of CAV data raises significant privacy concerns. Particularly, as such vehicles operate on public roads and collect information about the surroundings, they may pose privacy risks not only to the vehicle occupants but also to other road users.

Our findings indicate that the application of privacy patterns is challenging due to (i) inconsistent levels of abstraction of different patterns, (ii) lack of guidance for finding the right patterns for a given type of privacy threat, and (iii) gaps between the theoretical potential of privacy patterns and the practical complexities of their applicability in real-world use cases. Overall, the available patterns seem to address only a subset of the identified privacy threats. Thus, our work makes a significant contribution to a better understanding of what further research is needed to improve privacy engineering.

The remainder of the paper is structured as follows: First, we provide background and related work in Section 2 and details of our methodology in Section 3. We describe a representative robotaxi use case in Section 4, apply a LINDDUN-based analysis to identify privacy threats in Section 5, and analyze to what extent existing privacy patterns and our ideas for new privacy patterns help to find solutions for the identified privacy threats in Section 6. We discuss our findings in Section 7 and conclude in Section 8.

2. Background and Related Work

Privacy threat modeling is essential for privacy engineering, particularly to identify potential privacy threats. While there are several approaches to determine privacy threats, such as LINDDUN [14], LINDDUN GO [33],

[1] https://privacypatterns.org/

PANOPTIC™[2], and others [16], [29]—even in the automotive domain [28]—there is limited guidance on how to mitigate the threats. An earlier version of LINDDUN provided a taxonomy of mitigation strategies to provide a structured classification of common risk mitigation decisions, a means of selecting the appropriate strategies, and a list of privacy enhancing technologies (PETs).

Privacy strategies and *privacy patterns* were introduced by Hoepman [17] to generically solve privacy challenges. While the eight existing strategies are abstract concepts without any concrete implementation—classified as data-oriented (Minimize, Hide, Separate, Abstract) and process-oriented (Inform, Control, Enforce, Demonstrate) strategies—the patterns provide reusable solutions to common privacy problems. Later, *privacy tactics* by Colesky et al. [13] were added as an additional level of abstraction between privacy strategies and privacy patterns. Approaches of trying to trick users into disclosing personal data or giving consent against their real interest, are captured as *privacy dark patterns* [7], i. e., malicious patterns that deliberately weaken users' privacy. In the context of the Internet of Things, there has also been some work on the development [4] and application [25] of privacy patterns. Yet, the repository *privacypatterns.org* offers a wide range of generic patterns and has the potential to be used by many practitioners. For our research, we therefore analyzed all patterns from this collection. Undocumented patterns or patterns found elsewhere are not considered.

Caiza et al. [8], [9] empirically evaluated the application of 12 anonymity patterns in the design of a health monitoring system among students, and found that *applying and selecting privacy patterns* is difficult. Our work focuses on systematically investigating the effectiveness of a larger set of privacy patterns for automotive scenarios from practitioners and experts perspective. Further research on privacy patterns proposed extending patterns with properties [3] and architectural context information [10]–[12] to enhance their selection process. Other literature [5] links selecting privacy patterns with GDPR requirements [23].

While such literature focuses on improving the quality and the selection of patterns, the usefulness of privacy patterns in automotive systems such as robotaxi has not been systematically investigated. There is, however, related research on privacy in the general automotive domain. Bella et al. [6] investigate privacy policies for cars as well as user concerns and find that privacy for cars is insufficiently understood, mostly due to a lack of awareness. Syed et al. [30] propose a system model for enforcing purpose limitation, and Pape et al. [26] propose a system model to model and analyse suitable locations in the vehicle to add PETs. Concerning robotaxis, researchers proposed a privacy-preserving architecture [2] and others [21] examined people's perception of privacy in robotaxi which turned to be falsely positive. While understanding social perception of privacy in robotaxi is important, it is crucially necessary to study to what extent the available tools of privacy engineering such as threat modeling and privacy patterns are useful and practically applicable in scenarios like robotaxis in order to introduce a technically-sound privacy-preserving service.

3. Methodology

The main objective of this paper is to investigate the applicability and usefulness of existing privacy patterns when applied to the robotaxi scenario. For this purpose, we simulate a typical iterative design process of a robotaxi service, focusing on privacy. Specifically, we go through the following steps: (i) initial design, (ii) identification of privacy threats, (iii) attempting to improve the design by applying privacy patterns. Finally, we analyze and discuss our findings in terms of the applicability and usefulness of privacy patterns. In the following, we describe each step in more detail.

Initial design. We start with describing a typical robotaxi service, including the system model, the types of data processed, and relevant stakeholders, based on the relevant literature and existing similar services. We document our assumptions relating to the privacy properties of this service in particular detail. These assumptions are based on publicly known information about privacy practices in comparable services.

Identification of privacy threats. We apply the latest version of LINDDUN[3] to capture the privacy threats in the initial design. In particular, we use LINDDUN's privacy threat trees[4] for the threats of Linking, Identifying, Data Disclosure, Unawareness and Unintervenability, and Non-compliance. For each threat (i.e., each leaf of each tree), we assess its applicability to the different stakeholders in the robotaxi use case.

Applying privacy patterns. We use the privacy patterns as described on the *privacypatterns.org* website. For each of the threats identified in the previous step, we analyze if any of the available privacy patterns is applicable and would help mitigate the given threat. For the threats that could not be addressed by any of the existing patterns, we propose ideas for new privacy patterns.

4. Case Study: Robotaxi

We provide an overview of the system model, followed by an explicit description of privacy-related assumptions that will play an important role during the threat analysis.

4.1. System Model

The considered robotaxi system (see Figure 1) contains four parties: a *User* who wishes to order an autonomous vehicle from a Service Provider (*SP*) to transport a *Rider* from a pick-up location l_s at time t_s to a destination l_d at time t_d, and the original equipment manufacturer (*OEM*) of the vehicle. *User* and *Rider* are different roles that may be played either by the same person (if that person orders the ride for themselves) or by different persons (e.g., a hotel receptionist ordering a robotaxi for a guest of the hotel). In addition to these natural persons (*User* and *Rider*) and legal entities (*SP* and *OEM*), the *Vehicle* can be seen as a fifth party in the system.

To enable *User* to order a robotaxi (*Vehicle*) for *Rider*, *User* is required to have a user account with *SP*. To create

[2] https://ptmworkshop.gitlab.io/#/panoptic

[3] https://linddun.org/

[4] https://linddun.org/threat-trees/

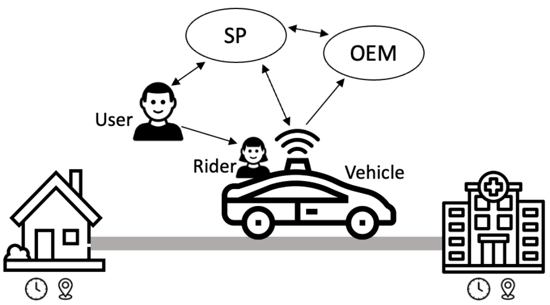

Figure 1: Basic system model.

TABLE 1: Data available to different parties.

Party	Available Data
User	Real-time location of ordered vehicle (and close-by vehicles to select from during ordering), routes, pickup-up and drop-off time, rider's information, complete ride history (rider names, locations, times, distances, routes, fares).
Rider	Number plate of robotaxi, locations and times, route, fare.
Vehicle	Rider authentication/ID, pick-up and drop-off locations and times of own riders, routes (times and locations), distances, fares, sensor data (riders height, weight), in-vehicle video, health monitoring, biometric data (e.g., fingerprints, DNA traces) of riders.
Service Provider (*SP*)	User account data (name, address, email, phone number, date of birth, payment information), pick-up and drop-off locations incl. times of all riders, history of past rides (including locations, times, distances, routes, fares), authentication material (tokens, biometric data), in-vehicle sensor data (e.g., seat pressure, health information, video feed).
OEM	Access to all sensor data (incl. passenger height and weight), camera data (inside/outside), location at specific time (routes), door events (trip start/end), health data.

the user account, *User* must provide at least a name, a valid phone number, and a valid means of payment (e.g., credit card information). To order a ride, *User* sends a request including *Rider*'s name(s), exact pick-up location(s) and time(s), and drop-off location(s) to *SP*. The *SP* processes the ride request and matches an available *Vehicle* with *Rider* based on their location(s), desired pick-up time, and possibly some vehicle preferences (e.g., vehicle class or interior). *SP* also handles the payment and offers special services such as a loyalty program for which the ride histories are stored. In addition, *SP* uses the collected data for fleet management, shares data with *OEM* to improve services, and might sell some of the collected data to third parties. To allow the ride matching, *Vehicle*s continually send their locations to *SP*. Once matched to *Rider*, *Vehicle* receives from *SP Rider*'s name(s), some rider authentication information (e.g., PIN or biometric data), pick-up and possibly drop-off locations, pick-up time, and *Rider*'s phone number. While driving, *Vehicle* collects different data, such as telemetry data, outside camera feed, lidar and radar signals, and possibly in-vehicle monitoring, such as in-vehicle camera feed (e.g., for abuse detection), health monitoring, and sensor data (e.g., seat pressure). In addition, *Vehicle* and *SP* learn the exact drop-off location(s) and time(s), the route driven, and the fare at the end of the ride. The *OEM* collects and analyzes telemetry data, e.g., for predictive maintenance, from its *Vehicle*s and has over-the-air access to all sensor data and CAN messages. Table 1 summarizes the data available to each party.

4.2. Assumptions

To clarify privacy-related aspects of the system design, we provide the following assumptions. These assumptions are used in Section 5 to enable better comprehension of our privacy threat analysis by linking one or more corresponding assumptions to the identified risks in Table 2a.

A1. The service requires a user registration with a unique identifier and identifying information (e.g., name, address, date of birth, payment information).

A2. *SP* and *OEM* share data, primarily for vehicle maintenance (e.g., vehicle diagnostics, logs). The shared data are used for debugging or improving services, and can contain *User*'s and *Rider*'s information such as *User*'s device specifications (e.g., operating system) and/or an identifier that is unique (globally or locally).

A3. *SP* has access to the (fine-grained) vehicle location traces (identifying Points of Interest (PoI)), in-vehicle sensor data (e.g., seat pressure, health information) including in-vehicle video feed (e.g., for abuse detection) of the *Vehicle*, the detailed ride histories (locations and times), routes, fares, number of passengers, and detailed mobility profiles of the *Rider*s. This allows to differentiate between rides and between *Rider*s of a *User*. This information is also used by *SP* for profiling *User* and *Rider*.

A4. *User* must provide the *Rider*s' name(s) to *SP*. *User* forwards *Rider* an authentication token (e.g., PIN, QR-code) for *Vehicle* from *SP*.

A5. Metadata is present in transmitted data streams (*User* to *SP* to *OEM*—following Assumption A2—and *Vehicle* to *OEM*) and can possibly identify *User*. This metadata might contain a device identifier, IP addresses, timestamps.

A6. The *OEM* has access to all vehicle sensors and CAN messages. This includes for instance camera stream (inside/outside), locations at specific times (routes), door and trunk events (trip start/end), and health data.

A7. *User* can track the assigned vehicle, and thus, is able to track *Rider* (via the application) during the use of the service. In addition, *User* has access to their ride history (i.e., rider names, pick-up and drop-off locations and times, routes, and fares).

A8. *SP* processes data about *User* for secondary purposes without *User* being aware of this. Such secondary purposes may include publishing research datasets[5], selling the data to third-party data monetization services that use it for targeted advertising, or making the data available to governmental agencies.

A9. *OEM* processes data about *Rider* for secondary purposes without *Rider* being aware of this. Such secondary purposes may include improving the design of future cars, or making the data available to governmental agencies.

A10. The data that *User* provides about *Rider* to *SP* is processed by *SP* for secondary purposes without User being aware of this. Such secondary purposes may include improving *SP*'s services, or making the data available to governmental agencies. Also, *Rider* is not aware of what information is shared by *User*.

A11. *SP* uses privacy dark patterns to make it difficult for *Users* to understand and set privacy-friendly preferences in their profiles.

[5] For example, an improperly anonymized New York City taxi dataset: https://blogs.lse.ac.uk/impactofsocialsciences/2014/07/16/nyc-improperly-anonymized-taxi-logs-pandurangan/

A12. *Users* have no means to access, rectify, or delete the data about them being processed at the *OEM*.
A13. The *OEM* is not informing *Rider* about the processing of their data and is not giving them the possibility to control (e.g., access, rectify, or erase) their data.

5. Privacy Threat Analysis

As previously mentioned, we employ the LINDDUN privacy threat modeling framework to identify potential privacy concerns in our scenario We consider all parties to behave in an honest-but-curious manner. We assume that *SP* would like to build a reputation as a privacy-friendly service provider, offering a privacy-preserving robotaxi service. *SP* uses the subsequent privacy analysis to guide its system design. We also assume that the knowledge that someone uses an ordinary taxi service every now and then does not reveal any further sensitive details. Therefore, our analysis excludes the threat type of Detecting, i.e., deducing the involvement of an individual through observation of communication, (trans)action side effects, or system responses. The threat type of Non-repudiation (i.e., the ability to attribute a claim to an individual) is also excluded because, according to [32], it is a rare privacy threat category, which only applies in niche applications such as e-voting and whistle-blower systems. Furthermore, the possibilities for mitigating non-repudiation threats in real-world scenarios are limited, especially, e.g., when billing information is strictly required.

In the following, we go through the remaining LINDDUN threat types, give their definition from LINDDUN, and analyze their relevance in the considered robotaxi scenario. The results are summarized in Table 2a.

Linking. ("Associating data items or interactions to learn more about an (unidentified) individual or group.") In our scenario, this could be linking a rider with a location to infer/learn something sensitive about the person. This threat is primarily executed by *SP* (not surprising, as it offers the service and collects data) or *OEM* (more surprising because it is not obvious why this would be needed for its operation). Note that *User* can threaten to link through derivation or inference, which is especially problematic when *User* and *Rider* only have a transactional relationship (e. g., *User* is a hotel concierge ordering a robotaxi for their client).

Identifying. ("Learning the identity of an individual.") In our scenario, the *SP* needs to know the *Rider* in order to offer proper services. Because of the data stored, the *OEM* and *User* can identify the *Rider*.

Data Disclosure. ("Excessively collecting, storing, processing or sharing personal data.") Major factors behind DD threats are type and granularity of the collected data, excessive data access, and further data processing by *SP* and *OEM*. In addition, data is shared with third parties.

Unawareness, Unintervenability. ("Insufficiently informing, involving or empowering individuals in the processing of personal data.") These threats only stem from the *SP* and the *OEM*.

Non-compliance. ("Deviating from security and data management best practices, standards, and legislation. The lack of adherence to legislation, regulation, standards, and best practices may lead to incomplete management of risks.") We decided not to include this category of threats in Table 2a, since determining the existence of these threats would require a legal discussion and divert our attention from the focus of this paper. Furthermore, to answer legal questions, we would need to decide which regulations to comply with and possibly resolve some conflicts (e. g., regulation for robotaxis might require the storage of some log files). Naturally, this heavily depends on the (geographical) area of legislation.

Takeaways

From Table 2a, we observe that the *SP* can perpetrate 30 out of 30 threats. This makes sense as *SP* provides the service and handles most of the dataflows. The *OEM* can perform 26 out of 30 threats. This is more surprising as the need for *OEM* to get access to personal data for its operation is not obvious. The *User* can act on 13 out of 30 threats. The *Rider* is not a threat actor at all. This analysis highlights a power imbalance among the actors. The *SP* plays a central role in this scenario, so should receive the most scrutiny. Also the *OEM* needs special attention because of the data sharing between *SP* and *OEM*, and because *OEM* has direct vehicle access (Assumption A6). In the next section, we investigate the privacy patterns that could help mitigate the identified threats.

6. Privacy Patterns Analysis

We have examined all patterns from *privacypatterns.org*, analyzed their applicability to the identified threats for our use case, and selected up to three patterns per threat that may be applied (individually or in combination) to help mitigate or even resolve a specific threat. In some cases, additional or different patterns could potentially be used to mitigate a particular threat; however, the objective of this analysis is not completeness, but rather the identification of gaps in the pattern landscape (see Section 3).

Applying existing patterns. We identified the following patterns as useful for mitigating some of the identified threats and list them here along with their definition:
P1. Attribute based credentials. Attribute Based Credentials (ABC) are an authentication mechanism that flexibly and selectively authenticates different attributes about an entity without revealing additional information about the entity (zero-knowledge property).
P2. Pseudo identity. Hide the identity by using a pseudonym and ensure a pseudonymous identity that can not be linked with a real identity during online interactions.
P3. Location granularity. Support minimization of data collection and distribution, e.g., by using location data with lower precision. Important when a service is collecting location data from or about a user, or transmitting location data about a user to a third-party.
P4. Added-noise measurement obfuscation. Add some noise to service operation measurements.
P5. Use of dummies. This pattern hides the actions taken by a user by adding fake actions that are indistinguishable from real actions.
P6. Aggregation gateway. Encrypt, aggregate, and decrypt at different places using homomorphic encryption.

TABLE 2: Threats based on LINDDUN for Linking, Identifying, Data Disclosure, and Unawareness. Table 2a shows the threats in the original scenario. In the column of an actor, ●^{Ax} means that the actor is threatening someone else's privacy based on assumption Ax. Table 2b shows the mitigated (or resolved) threats ◑^{Px} after applying existing privacy pattern(s) (Px) and the remaining privacy threats ⊙^{Sx} that could be addressed by introducing a new pattern (Sx).

(a) For Original Scenario

	Threat	SP	OEM	User	Rider
Linking (L)	*Linking through identifiers*				
	1.1 Unique identifier	●^{A1}	●^{A2}	○	○
	Linking through combination				
	2.1.1 Single user	●^{A3}	●^{A6}	○	○
	2.1.2 Multiple users	●^{A3}	●^{A6}	○	○
	Linking through derivation or inference (profiling)				
	2.2.1 Individual	●^{A3}	●^{A6}	●^{A7}	○
	2.2.2 Group	●^{A3}	●^{A6}	●^{A7}	○
	2.2.3 (Dis)similarity	●^{A3}	●^{A6}	○	○
Identifying (I)	*Identified information*				
	1.1 Identified info	●^{A1,A3,A4}	○	○	○
	1.2 Metadata	●^{A5}	●^{A5,A6}	○	○
	Identifiable information				
	2.1.1 Identifier	●^{A1}	○	○	○
	2.1.2 Quasi-identifier	●^{A3}	●^{A6}	●^{A7}	○
	2.2 Attributes	●^{A3}	●^{A6}	●^{A4,A7}	○
	2.3 Distinguishability	●^{A3}	●^{A6}	●^{A7}	○
Data Disclosure (DD)	*Unnecessary data types (DT)*				
	1.1 DT sensitivity	●^{A3}	●^{A6}	●^{A7}	○
	1.2 DT granularity	●^{A1,A3}	●^{A6}	●^{A7}	○
	1.3 DT encoding	●^{A5}	●^{A6}	○	○
	Excessive volume				
	2.1 Amount	●^{A3}	●^{A6}	●^{A4}	○
	2.2 Frequency	●^{A3}	●^{A6}	●^{A7}	○
	2.3 Data subjects	●^{A4}	●^{A6}	●^{A4}	○
	Unnecessary processing				
	3.1 Treatment	●^{A3}	●^{A6}	○	○
	3.2 Propagation	●^{A8}	●^{A9}	○	○
	3.3 Implicit disclosure	●^{A3}	●^{A6}	●^{A7}	○
	3.4 Duration/retention	●^{A2}	●^{A2,A6}	●^{A7}	○
	Involved parties and exposure				
	4.1.1 Fixed parties	●^{A8}	●^{A9}	○	○
	4.1.2 Dynamic parties	●^{A8}	●^{A9}	○	○
	4.2 Data accessibility	●^{A8}	○	○	○
Unawaren. (U)	*Unawareness of processing*				
	1.1 As data subject	●^{A8,A11}	●^{A13}	○	○
	1.2 As a user sharing	●^{A10,A11}	●^{A13}	○	○
	Lack of data subject control				
	2.1 Preferences	●^{A12}	●^{A13}	○	○
	2.2 Access	●^{A12}	●^{A13}	○	○
	2.3 Rectification/erasure	●^{A12}	●^{A13}	○	○
Sum	30 threats	●:30 ○: 0	●:27 ○: 3	●:12 ○:18	●: 0 ○:30

(b) For Scenario with Privacy Patterns Applied

	Threat	SP	OEM	User	Rider
Linking (L)	*Linking through identifiers*				
	1.1 Unique identifier	◑^{P1}	◑^{P2}	○	○
	Linking through combination				
	2.1.1 Single user	◑^{P3,P4}	◑^{P3,P4}	○	○
	2.1.2 Multiple users	◑^{P3,P4,P5}	◑^{P3,P4}	○	○
	Linking through derivation or inference (profiling)				
	2.2.1 Individual	◑^{P2,P3,P4}	⊙^{S1}	◑^{P3}	○
	2.2.2 Group	◑^{P3,P4,P5}	⊙^{S1}	◑^{P3}	○
	2.2.3 (Dis)similarity	◑^{P3,P6,P7}	⊙^{S1}	○	○
Identifying (I)	*Identified information*				
	1.1 Identified info	◑^{P1,P3}	○	○	○
	1.2 Metadata	◑^{P8,P9}	◑^{P3,P8}	○	○
	Identifiable information				
	2.1.1 Identifier	◑^{P2}	○	○	○
	2.1.2 Quasi-identifier	◑^{P3,P4}	◑^{P3,P4}	◑^{P3,P4}	○
	2.2 Attributes	◑^{P1,P3,P4}	◑^{P3,P4}	◑^{P2,P3}	○
	2.3 Distinguishability	◑^{P1,P3,P10}	◑^{P3,P4}	◑^{P2,P3}	○
Data Disclosure (DD)	*Unnecessary data types (DT)*				
	1.1 DT sensitivity	⊙^{S2}	⊙^{S2}	⊙^{S2}	○
	1.2 DT granularity	⊙^{S4}	⊙^{S4}	⊙^{S4}	○
	1.3 DT encoding	⊙^{S2}	⊙^{S2}	○	○
	Excessive volume				
	2.1 Amount	⊙^{S2,S3}	⊙^{S2,S3}	⊙^{S2}	○
	2.2 Frequency	⊙^{S2,S3}	⊙^{S2,S3}	⊙^{S2}	○
	2.3 Data subjects	⊙^{S2,S3}	⊙^{S2,S3}	⊙^{S2}	○
	Unnecessary processing				
	3.1 Treatment	⊙^{S3}	⊙^{S3}	○	○
	3.2 Propagation	⊙^{S3}	⊙^{S3}	○	○
	3.3 Implicit disclosure	⊙^{S3}	⊙^{S3}	⊙^{S3}	○
	3.4 Duration/retention	⊙^{S5}	⊙^{S5}	⊙^{S5}	○
	Involved parties and exposure				
	4.1.1 Fixed parties	⊙^{S6}	⊙^{S6}	○	○
	4.1.2 Dynamic parties	⊙^{S6}	⊙^{S6}	○	○
	4.2 Data accessibility	⊙^{S6}	○	○	○
Unawaren. (U)	*Unawareness of processing*				
	1.1 As data subject	◑^{P11}	◑^{P12}	○	○
	1.2 As a user sharing	◑^{P12}	◑^{P12}	○	○
	Lack of data subject control				
	2.1 Preferences	◑^{P13}	◑^{P13}	○	○
	2.2 Access	◑^{P14}	◑^{P14}	○	○
	2.3 Rectification/erasure	⊙^{S7}	⊙^{S7}	○	○
Sum	30 threats	◑:16 ⊙:14	◑:11 ⊙:16	◑:5 ⊙:7	◑:0 ⊙:0

P7. Trustworthy privacy plug-in. Aggregate usage records at the user side in a trustworthy manner before transmitting the data to the service provider.

P8. Strip metadata. Strip potentially sensitive metadata that isn't directly visible to the end user.

P9. Onion routing. This pattern provides unlinkability between senders and receivers by encapsulating the data in different layers of encryption, limiting the knowledge of each node along the delivery path.

P10. Anonymity set. Aggregate multiple entities into a set, such that they cannot be distinguished anymore.

P11. Minimal information asymmetry. Prevent users from being disenfranchised by their lack of familiarity with the policies, potential risks, and their agency within processing.

P12. Privacy icons (notice/labels). A privacy policy which is hard to understand by general audience is summarized and translated into commonly agreed visual icons. A privacy icon is worth a thousand-word policy.

P13. Platform for privacy preferences. Use privacy policies which consist of standardized and extensible vocabulary and data element sets, both of which user agents should be aware of, to streamline their review by eliminating redundancies.

P14. Reasonable Level of Control. Let users share selectively (push) and make available (pull) specific information to predefined groups or individuals.

Ideas for new patterns. Since some of the threats are not addressable with the current privacy patterns, we provide a list of *ideas for new privacy patterns* that could help mitigate the remaining threats along with the strategies [17] these ideas fall under. Recall that the aim of this paper is merely to identify gaps in the pattern landscape and not to provide new, fully-fledged privacy patterns.

S1. Secure data processing. Technologies such as (fully) homomorphic encryption, secure multi-party computation, or confidential computing using trusted execution environments allow data processing on encrypted data without having access to the plaintext data. Secure function evaluation provides privacy and confidentiality of the input data. Such a pattern would be located in the Hide strategy.

S2. Reduce data collection. This idea tackles reducing the collection of data to the strictly required data only. If data is not required for a particular service, it should not be collected. A pattern that addresses this idea falls under the Minimize strategy.

S3. Reduce data processing. This refers to reducing the processing frequency, amount of data, and the involved individuals to the strictly required. A pattern that addresses this idea falls under the Minimize strategy.

S4. Reduce data granularity. This idea refers to adjust the granularity of data to the necessary level of precision. Proof of age, for example, does not always require the exact date of birth, but merely a bit of information as to whether the user is above or below a certain age threshold. This pattern idea is related to the already existing Location granularity pattern. A pattern that addresses this idea falls under the Abstract strategy.

S5. Limit Data retention. This idea refers to deleting user data if it is not needed anymore, or after a certain (usually pre-defined) amount of time. Particularly, data that is no longer required or that violates retention policies (based on context, content, or purpose) should be automatically deleted to reduce the risk of attacks or misuse. A pattern that addresses this idea falls under the Minimize strategy.

S6. Limit third party access. Restrict third party data access to the minimum. Many *SP*s use third-party (tracking) services or sell user data to third-party data monetization services. This should be restricted to a minimum, as each additional service poses a greater threat of data disclosure or misuse. A pattern that addresses this idea falls under the Minimize strategy.

S7. Provide support for requests to delete and rectify data. Many *SP*s do not provide an option for the users to rectify and permanently delete data relating to them or have stored data erased. This could be implemented in an app or web interface addressing the Control strategy.

Takeaways

As shown in Table 2b, existing privacy patterns could help mitigate 16 threats at *SP*, 11 at *OEM*, and five at *User*. In particular, these patterns help mitigate all of the Identifiability threats (using 6 Px), 78% of the Linking (using 7 Px), and 80% of the Unawareness ones (using 8 Px). We observe that $P3$, $P4$ and $P2$ are the most prevalent patterns. On the other hand, the suggested patterns (Sx) can help mitigate all the remaining threats. Especially, $S1$ mitigates the 22% remaining Linking threats, five Sx cover all the Data Disclosure threats, and two Sx answer the remaining 20% of the Unawareness threats. We note that $S2$, $S3$ and $S6$ are the most prevalent Sx, and are under the Minimize strategy. We also observe that, contrary to Px, when a Sx mitigates a threat at one stakeholder, it tends to do so for all the others as well.

7. Discussion

In our exploration of the application of privacy patterns within the automotive industry, with a focus on the robotaxi scenario, we uncovered a number of challenges and opportunities. At first glance, our assessment suggests that privacy patterns can offer robust solutions to privacy threats. However, considering real-world requirements and scenario specifics, the practical application of such patterns has a variety of implications that challenge our initial optimism. These challenges fall into three areas: application of patterns, quality and coverage of patterns, and evaluating the impact of applying patterns.

7.1. Application of Privacy Patterns

The application of privacy patterns is challenging, as significant effort may be needed to implement privacy patterns [8]. We describe the difficulties faced below.

Conflicting requirements. Firstly, the direct application of privacy patterns often encounters obstacles due to conflicting demands between preserving user privacy and maintaining essential system functionality and utility [20]. For example, a robotaxi necessitates and relies on precise location data (e. g., pick-up location) to operate effectively, presenting a fundamental clash with privacy patterns such as *location granularity* for the specific purpose of pick-up and possibly drop-off. However, the pattern may be relevant for further backend processing of location data. Similarly, external and non-functional requirements, such as the legal obligation to retain billing information for tax purposes, further complicate the straightforward application of privacy patterns that would enable anonymity or pseudonymity. As these examples show, deciding whether and, if yes, how exactly a pattern is applicable can be challenging in practice.

Specificity and composition of patterns. The application of privacy patterns also suffers from a lack of specificity in guiding the realization of a privacy solution to mitigate given privacy threats. Privacy patterns provide a conceptual framework for addressing privacy threats and they are used during the design phase of the system life cycle [18]. However, they fall short in offering detailed methodologies and know-how for their application. In our investigations, it became evident that the effective mitigation of a certain privacy threat frequently requires a combination of multiple patterns that together could provide a holistic privacy solution to that threat. However, no guidance is available for composing multiple privacy patterns. Furthermore, even if suitable privacy patterns are identified, selecting a suitable privacy enhancing technology (PET) is still hard [19]. This highlights a significant gap between the theoretical potential of privacy patterns and the practical complexities of their applicability in real-world cases.

Real-world application. Furthermore, certain privacy threats particular to the robotaxi scenario prove intrinsically difficult to address. Events that are highly distinctive by nature can evidently reveal information about the user, despite applying privacy patterns. For instance, ordering a robotaxi in an isolated area will render the pattern location granularity useless as this is a distinguishing event. Such scenarios highlight the limitations of current privacy-preserving approaches and shed light on the need for innovative solutions that can accommodate the unpredictability of real-world scenarios' context.

7.2. Quality and coverage of privacy patterns

Previous work [8], [9] highlighted some issues regarding the collection, maintenance, and quality assurance of privacy patterns. One takeaway is that normalizing patterns and reviewing them before publication is important. Additionally, the authors found that patterns are described using different terminology and writing styles. This is in line with our findings, as described below.

Pattern names. The names of patterns are often too generic (e.g., for data disclosure). This makes finding the right pattern harder for those not already familiar with the pattern databases.

Missing patterns. As shown in Section 6, our investigation identified certain gaps in the current landscape of privacy patterns. Particularly, there is an absence of patterns addressing critical aspects, such as *data deletion* or *replacing sensitive data*, which are vital for ensuring privacy over time. This gap highlights the need for a more comprehensive set of privacy patterns, as well as for methods to check the completeness of the set of available patterns. Some patterns could be tailored to the specific challenges of the automotive sector, e.g., patterns to address special user interfaces for automotive use cases.

Distribution of patterns. As of now, there is a large imbalance between the number of patterns per privacy design strategy. There are 59 process-oriented patterns (Inform: 33, Control: 22, Enforce: 4, Demonstrate: 0) and only 19 data-oriented patterns (Minimize: 4, Hide: 11, Separate: 3, Abstract: 1). This might be a problem since process-oriented patterns tend to primarily solve compliance issues around privacy, while data-oriented patterns tend to improve the used data from a privacy perspective, e. g., by minimizing it. Our pattern ideas contribute to a more balanced distribution as we mostly suggested data-oriented privacy patterns (six out of seven) to solve the challenges of our use case.

Managing patterns. When solving a privacy problem, it is not easy to generalize the solution to a new pattern. It creates additional effort after the problem is solved, the pattern would need to be fully formulated, and added to a pattern collection (like *privacypatterns.org*). There needs to be some quality assurance and the pattern needs to be maintained. A first starting point to resolve this issue might be to get inspiration from software documentation where there are similar issues [1].

Tailoring privacy patterns. Often, the identified patterns only provide a first idea on how exactly to tackle the threat. These patterns then need to be tailored to the specific use case. However, tailoring a pattern might be challenging. It could also raise the question of when tailoring an existing privacy pattern would result in a new pattern.

Pattern goals. Pattern goals are not explicitly stated. Some patterns, e.g., from Inform strategy, exist to support compliance, whereas some target technical solutions, e.g., from Minimize strategy (cf. Distribution of patterns).

7.3. Privacy Evaluation and Goals of Services

Caiza et al. [9] pointed out the need for mechanisms to evaluate the impact of applying privacy patterns. This holds for both experimental setup as in their study, as well as for real world applications of privacy patterns.

Privacy versus consent theater. The goal of the privacy enhancements is not clearly defined. In some cases, companies dealing with data are afraid of being incompliant with privacy regulations. They see the regulations as a hurdle that needs to be overcome instead of a baseline that represents the absolute minimum for privacy. Thus, their main goal might not be to offer a privacy-friendly service but to just be compliant. It has also been shown that requesting the users' consent might help in being compliant, but, in most cases, does not help to achieve privacy-friendly services [15]. Accordingly and as we discussed earlier, the vast majority of the current privacy patterns help in compliance-related matters, meaning that their deployment may not fully enhance privacy in the systems.

Minimal data needed for service. Even if the goal is to offer a privacy-friendly service, it is often not clear what is the required minimum amount of data to offer services [31]. Furthermore, besides legally required retention periods, it is not always clear when the data is not "needed" anymore and can be deleted.

Privacy level assessment. To assess the benefits of applying a privacy pattern, there is a need to assess the privacy level of a service. However, this evaluation requires agreement on the risk levels associated with each privacy threat. This is where Privacy Impact Assessments could complement LINDDUN-type methodologies.

8. Conclusion and Future Work

Privacy patterns hold a promise for enhancing privacy in automotive applications. This paper explored the usefulness of privacy patterns in improving privacy in future automotive systems. In particular, we considered a robotaxi scenario and investigated the applicability of privacy patterns along with their capabilities to solve LINDDUN privacy threats. We identified challenges in applying existing privacy patterns and proposed new ideas for missing ones. Our analysis showed that the practical implementation of the patterns is challenging due to (i) conflicts between functionality and external requirements, (ii) existing gaps in the pattern landscape, and (iii) a lack of detailed implementation guidelines. Thus, significant effort is needed to bridge the gap between the theoretical potential of privacy patterns and their practical applicability in real-world systems.

Our work opens the door to various areas for future work. Existing privacy patterns focus on the digital world, so we took a similar focus. However, since vehicles are part of the physical world as well, future work could investigate to what extent patterns already consider the physical aspect of cyber-physical systems, and how to adapt them. In practice, a Mobility-as-a-Service scenario exhibits additional stakeholders, e.g., payment providers and other service providers, resulting in additional privacy threats. Future work should, therefore, consider a holistic set of stakeholders and examine the usefulness of privacy patterns in such cases. Furthermore, privacy threats stemming from automotive sensors are an under-explored research area because of the lack of a clear mapping of sensor data to (potential) personally identifiable information. Currently, existing privacy risk assessments for automotive sensors do not consider the application of patterns to address identified risks [27]. Lastly, while we did not identify the existence or need for automotive-specific patterns, more research is needed to explore whether domain-specific patterns, for the automotive or other domains, would be useful.

Acknowledgements

This work was inspired by privacy engineering discussions at Dagstuhl Seminar 23242 ”Privacy Protection of Automated and Self-Driving Vehicles”. The work was supported in part by the German Federal Ministry of Education and Research (BMBF) under grant number 16KIS1382, and by the U.S. National Science Foundation (NSF) under grant number 2245323. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the BMBF or NSF.

References

[1] Emad Aghajani, Csaba Nagy, Mario Linares-Vásquez, Laura Moreno, Gabriele Bavota, Michele Lanza, and David C Shepherd. Software documentation: the practitioners’ perspective. In *International Conference on Software Engineering*, 2020.

[2] Ala’a Al-Momani, Frank Kargl, Robert Schmidt, and Christoph Bösch. iRide: A privacy-preserving architecture for self-driving cabs service. In *Vehicular Networking Conference*. IEEE, 2018.

[3] Ala’a Al-Momani, Kim Wuyts, Laurens Sion, Frank Kargl, Wouter Joosen, Benjamin Erb, and Christoph Bösch. Land of the lost: privacy patterns’ forgotten properties: enhancing selection-support for privacy patterns. In *ACM SAC*, 2021.

[4] Nada Alhirabi, Stephanie Beaumont, Omer F. Rana, and Charith Perera. Privacy-patterns for IoT application developers. In *Adjunct Proceedings of UbiComp/ISWC*, 2022.

[5] Vita Santa Barletta, Giuseppe Desolda, Domenico Gigante, Rosa Lanzilotti, and Marco Saltarella. From GDPR to privacy design patterns: The MATERIALIST framework. In *SECRYPT*, 2022.

[6] Giampaolo Bella, Pietro Biondi, Marco De Vincenzi, and Giuseppe Tudisco. Privacy and modern cars through a dual lens. In *EuroS&PW*, 2021.

[7] Christoph Bösch, Benjamin Erb, Frank Kargl, Henning Kopp, and Stefan Pfattheicher. Tales from the dark side: Privacy dark strategies and privacy dark patterns. *Privacy Enhancing Technologies*, 2016.

[8] Julio C. Caiza, José M. del Álamo, and Danny S. Guamán. A framework and roadmap for enhancing the application of privacy design patterns. In *ACM SAC*, 2020.

[9] Julio C. Caiza, José M. del Álamo, Danny S. Guamán, and Angel Jaramillo-Alcázar. An exploratory experiment on privacy patterns: limitations and possibilities. In *ACM SAC*, 2021.

[10] Su Yen Chia, Xiwei Xu, Ming Ding, David B. Smith, Hye-Young Paik, and Liming Zhu. A selection model of privacy patterns. In *IEEE ICSA*, 2023.

[11] Su Yen Chia, Xiwei Xu, Hye-Young Paik, and Liming Zhu. Analysing and extending privacy patterns with architectural context. In *ACM SAC*, 2021.

[12] Su Yen Chia, Xiwei Xu, Hye-Young Paik, and Liming Zhu. Analysis of privacy patterns from an architectural perspective. In *ICSA Companion*, 2022.

[13] Michael Colesky, Jaap-Henk Hoepman, and Christiaan Hillen. A critical analysis of privacy design strategies. In *SPW*, 2016.

[14] Mina Deng, Kim Wuyts, Riccardo Scandariato, Bart Preneel, and Wouter Joosen. A privacy threat analysis framework: supporting the elicitation and fulfillment of privacy requirements. *Requirements Engineering*, 16(1):3–32, 2011.

[15] Matthias Fassl, Lea Theresa Gröber, and Katharina Krombholz. Stop the consent theater. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 2021.

[16] Rafa Galvez and Seda Gurses. The odyssey: Modeling privacy threats in a brave new world. In *EuroS&PW*, 2018.

[17] Jaap-Henk Hoepman. Privacy design strategies - (extended abstract). In *ICT Systems Security and Privacy Protection*. Springer, 2014.

[18] Jaap-Henk Hoepman. Privacy design strategies (the little blue book). Radboud University, 2018.

[19] Sascha Löbner, Frédéric Tronnier, Sebastian Pape, and Kai Rannenberg. Comparison of de-identification techniques for priv. preserving data analysis in vehicular data sharing. In *ACM CSCS*, 2021.

[20] Zoltán Ádám Mann, Jonathan Petit, Sarah M. Thornton, Michael Buchholz, and Jason Millar. SPIDER: Interplay assessment method for privacy and other values. In *EuroS&PW*, 2024.

[21] Johanna Meurer, Christina Pakusch, Gunnar Stevens, Dave Randall, and Volker Wulf. A wizard of oz study on passengers’ experiences of a robo-taxi service in real-life settings. In *ACM Designing Interactive Systems Conference*, 2020.

[22] Milos Mladenovic. Mobility as a service. In *International Encyclopedia of Transportation*. Academic Press, 2021.

[23] Council of the European Union. Regulation (EU) 2016/679 (General Data Protection Regulation). Official Journal of the European Union vol. 59, 2016.

[24] Alkis Papadoulis, Mohammed Quddus, and Marianna Imprialou. Evaluating the safety impact of connected and autonomous vehicles on motorways. *Accident Analysis & Prevention*, 2019.

[25] Sebastian Pape and Kai Rannenberg. Applying privacy patterns to the Internet of Things’ (IoT) architecture. *Mobile Networks and Applications*, 24(3):925–933, 2019.

[26] Sebastian Pape, Sarah Syed-Winkler, Armando Miguel Garcia, Badreddine Chah, Anis Bkakria, Matthias Hiller, Tobias Walcher, Alexandre Lombard, Abdeljalil Abbas-Turki, and Reda Yaich. A systematic approach for automotive privacy management. In *ACM CSCS*, 2023.

[27] Mert D Pesé and Kang G Shin. Survey of automotive privacy regulations and privacy-related attacks. SAE Technical Paper 2019-01-0479, 2019.

[28] Mario Raciti and Giampaolo Bella. A threat model for soft privacy on smart cars. In *EuroS&PW*, 2023.

[29] Laurens Sion, Kim Wuyts, Koen Yskout, Dimitri Van Landuyt, and Wouter Joosen. Interaction-based privacy threat elicitation. In *EuroS&PW*, 2018.

[30] Sarah Syed-Winkler, Sebastian Pape, and Ahmad Sabouri. A data protection-oriented system model enforcing purpose limitation for connected mobility. In *ACM CSCS*, 2022.

[31] Tjerk Timan and Zoltan Mann. Data protection in the era of artificial intelligence: Trends, existing solutions and recommendations for privacy-preserving technologies. In *The Elements of Big Data Value*, pages 153–175. 2021.

[32] Kim Wuyts. *Privacy threats in software architectures*. PhD thesis, KU Leuven, 2015.

[33] Kim Wuyts, Laurens Sion, and Wouter Joosen. LINDDUN GO: A lightweight approach to privacy threat modeling. In *EuroS&PW*, 2020.