Journal of Ubiquitous Systems & Pervasive Networks
Volume 19, No. 1 (2023) pp. 25-32

Exploring Privacy Threats in Connected and Autonomous Vehicles: An Analysis

Badreddine Chah [a]*, Alexandre Lombard [a], *Anis Bkakria* [b], *Reda Yaich* [b] *Abdeljalil Abbas-Turki* [a]

[a]*CIAD UMR 7533, Univ. Bourgogne-Franche-Comté, UTBM, F-90010 Belfort, France*
[b]*IRT SystemX, Palaiseau 91120, France*

Abstract

This paper conducts a privacy threat analysis of connected and autonomous vehicles to address the growing concerns regarding the privacy of sensitive information collected by these vehicles. By following the LINDDUN GO methodology, this paper aims to identify privacy risks in the overall connected and autonomous vehicles architecture based on formal privacy requirements. The proposed analysis focuses on a specific use case, assisting manufacturers in implementing privacy requirements and enhancing privacy protection. This research provides valuable insights into potential risks and vulnerabilities, contributing to the development of privacy-respecting connected and autonomous vehicle systems.

Keywords: *Privacy, Security, Threat Analysis, Connected and Autonomous Vehicle, Privacy engineering framework.*

1. Introduction

In today's world, numerous automotive companies are increasingly offering a variety of services based on the technological capabilities of their customers' vehicles. To provide these services, automotive companies require data on each client. This data is generally considered sensitive and can be collected, processed, and analyzed by companies without sufficient access controls. As vehicles become more connected and autonomous, the amount of data collected and transmitted increases, leading to a greater variety of services being provided. It causes significant privacy concerns in the market.

Nevertheless, to prevent possible abuses, the collection of personal data requires a strict framework. At the European Union (EU) level, the General Data Protection Regulation (GDPR) [1] serves as the legal regulation that safeguards data usage and protects users' privacy and personal information. The main objective of the GDPR is to ensure that information serves the customer without compromising human identity, rights, privacy, or individual and public freedoms. Similarly, other countries have dedicated organizations for data protection and privacy. The GDPR sets forth principles that must be strictly adhered to preserve personal data. These principles include:

- Customers have the right to determine the use made of their personal data.
- Concerned manufacturers must have a legal basis for the processing of people's data, following the conditions of section 6 of the GDPR [1].
- Lawfulness, fairness, and transparency: manufacturers need to process data in a clear, fair, and lawful manner.
- Purpose limitation: the purpose of the use of personal data must be deterministic, explicit, and legal.
- Data minimization: it is necessary to be able to control the collection of data needed for a specific purpose.
- Storage limitation: the retention of personal data must be limited unless the data are anonymized.
- Integrity and Confidentiality: it is necessary to take into consideration the securitization of the data circulating inside the vehicle, but also outside the vehicle.

The main challenge for companies lies in providing their services while adhering to the privacy requirements outlined by the GDPR. To address this concern, we propose conducting a privacy threat analysis that focuses on several key aspects: *Linkability*, *Identifiability*, *Non-repudiation*, *Detectability*, and *Disclosure of information*. These aspects are categorized under the LINDDUN methodologies.

*Corresponding author. Tel.: +33-384-583-319
Fax: +33-384-583-342; E-mail: Badreddine.chah@utbm.fr

DOI: 10.5383/JUSPN.19.01.003

In [2], the authors establish a robust correlation between the GDPR principles and the LINDDUN aspects by mapping them together. This correlation provides valuable insights into the relationship between the two frameworks. Considering that Connected Autonomous Vehicles (CAVs) are regarded as complex systems for studying privacy threats, addressing the aforementioned issue requires us to answer the following questions:

1. What type of data is collected and processed in the CAV system?
2. What are the privacy threats present in the CAV environment?
3. Which adversary models should be considered, and what is the potential damage to privacy?
4. What are the vulnerabilities and attack scenarios that could help define misuse case scenarios?
5. What are the privacy-enhancing techniques (PETs) that can mitigate the risk of privacy loss or damage?

In this article, we further extend our previous study [3] by conducting a privacy threat analysis, aiming to fully apply the LINDDUN methodology to CAVs while highlighting potential attacks before they occur. By conducting this analysis, our goal was to identify potential attacks before they happen, with a specific emphasis on data-driven privacy to ensure compliance with GDPR requirements. This work completes the missing elements that we could not address in the previous papers. The paper is structured as follows: in Section 2., we provide the necessary background information to clarify and prepare for understanding the rest of the paper. Then the Section 3, we present the core of our contribution, where we apply the new version of the LINDDUN GO methodology to CAV use cases. In Section 4., we delve into the analysis of our results and their placement within the existing literature. Subsequently, we conclude our work and outline the directions for future research.

2. Background

2.1. Data to Protect

As CAV technologies develop, the concerns over the privacy of the data generated, collected, and analyzed have become more voluminous. Among the types of data circulating in the vehicle, we aim to determine what types of data are considered personal data. The personal data concerned include all data linked to a physical entity (driver, vehicle owner, passenger). We select the types of data to protect that are relevant to our use case:

- **Vehicle/Owners ID:** Vehicles participating in the use case exchange identifying information to establish communication and coordination. This can include Vehicle serial number, last name of owner, owner name, owner address, owner telephone number, owner e-mail address, etc.
- **Position and speed:** Each vehicle shares its position and speed, in real-time, with the PSP and/or other vehicles on the road.
- **Vehicles data:** CAV may exchange sensor data, such as radar or LiDAR measurements, to enhance perception and situational awareness within the road. Plus, it can exchange Data related to the use of the vehicle by the driver or its occupants (e.g. data related to driving styles, mileage, life in the vehicle, etc.).
- **Control commands:** Some use cases involve coordinated movements, so control commands are exchanged to synchronize acceleration, braking, and steering between vehicles. These commands are critical to maintaining the desired service.
- **Communication integrity and security data:** To ensure safe and reliable communication, sensitive data related to communication integrity, authentication, encryption keys, and security protocols are exchanged. This protects against unauthorized access, manipulation, or malicious attacks on the use case system.

The sensitive information links to individuals who may not be willing to share it, mainly for the preservation of their privacy, in compliance with the GDPR principles.

2.2. Privacy Threat Modeling: LINDDUN GO

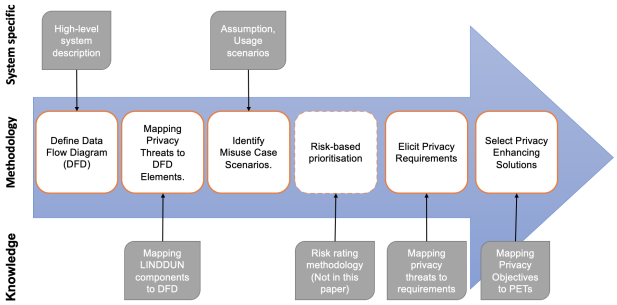

Fig. 1. The *LINDDUN* methodology and the required system-specific knowledge.

LINDDUN [4] is an acronym that represents the categories of threats in privacy analysis. These threats can be categorized as hard privacy threats, which include *Linkability*, *Identifiability*, *Non-repudiation*, *Detectability*, and *Disclosure of information*. There are also soft privacy threats, such as *Unawareness* and *Non-compliance*. The LINDDUN methodology serves as a privacy threat modeling framework that aids in systematically identifying and mitigating privacy threats.

The complete LINDDUN framework involves iteratively exploring all threat trees, which can be time-consuming due to the presence of over 100 leaf nodes. To streamline the process, a new version called LINDDUN GO [5] has been introduced. LINDDUN GO condenses the knowledge into 35 threat-type cards that need to be considered. This reduction is achieved by combining related threat types. For example, the lack of data portability is closely linked to the lack of data access. Furthermore, less significant threats are excluded from the LINDDUN GO methodology.

In this work, our exclusive focus is on hard privacy threats, which will be elaborated on in the subsequent section. We will delve into the aspects of *Linkability*, *Identifiability*, *Non-repudiation*, *Detectability*, and *Disclosure of information* to thoroughly analyze and address the relevant privacy concerns.

LINDDUN presents a threat modeling process consisting of six high-level steps, as depicted in Fig. 1. *Step 1: Define Data Flow Diagram (DFD).* DFD is created based on the high-level system description. In our case study, the analyzed CAV is decomposed

into relevant logical and structural elements, and for each of these parts, the corresponding threats are reviewed. *Step2: Mapping privacy threats to DFD Elements.* Each element of the DFD is subject to certain privacy threats. *Step 3: Identify misuse case scenarios.* We start by eliciting privacy threats, using threat-type cards. In this step, we elicit all threats related to CAVs. *Step 4: Risk-based prioritization.* This step assesses the severity of identified attack risks and makes informed decisions on how to address them. *Step 5: Elicit privacy requirements.* This defines the privacy objectives that need to be fulfilled. *Step 6: Select privacy-enhancing solutions.* This last step maps the privacy objectives to the PETs available in the literature.

It is important to note that Steps 5 and 6 are considered "white hat" activities focused on defining privacy requirements and selecting appropriate privacy-enhancing solutions. In the scope of this paper, we concentrate on expanding the content of our previous work [3]. Therefore, we will need to refer to elements from that paper. The new version of LINDDUN [5] aims to streamline the time-consuming threat analysis process by replacing traditional threat trees with threat card types, which will be explained below.

In the new version of LINDDUN, as described in [5], the authors introduce a collection of threat type cards that provide a description of potential privacy threats. These cards serve as an extension and structured representation of the threat type descriptions found in LINDDUN's threat tree, as presented in previous works [3, 4]. According to [5], each threat type card consistently emphasizes the following important key points:

- *Hotspots:* Indicate the specific area within the system where the threat is present.
- *Threat source:* Specifies the origin type of the threat, distinguishing between organizational, external to the system, or the receiving party involved in the interaction.
- *Summary*: This is a short description of the threat type.
- *Elicitation questions:* Two questions are employed to aid in determining the applicability of the threat type. The first question primarily assesses whether the prerequisites are met, while the second question helps evaluate the actual applicability of the threat.
- *Examples:* Give illustration(s) of the threat type.
- *Consequences/impact:* Outlines the potential impacts or consequences on privacy if the threat is successfully exploited, to rationale about the threat's importance.

As previously mentioned, steps 1, 2, 5, and 6 were presented in our previous work [3]. In the subsequent section, we take a bold step forward and present the necessary elements to analyze step 3, which focuses on identifying misuse case scenarios specific to CAVs based on threat-type cards. This endeavor aims to build upon the foundation laid in our previous work and provide a comprehensive understanding of potential misuse cases in the context of CAVs.

3. LINDDUN Application to CAV Based threat type Card

CAVs represent a transformative technology with immense potential. They offer numerous benefits, including the potential to reduce road accidents, enhance the quality of life, and improve the efficiency of transportation systems. A key aspect of CAVs

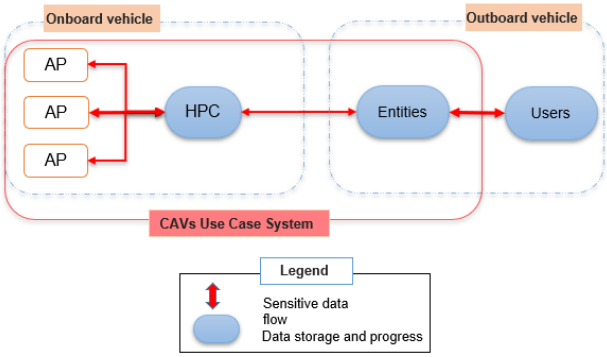

Fig. 2. The Data Flow Diagram (DFD) of CAV. AP: Access Point, HPC: high-performance computer.

is their communication capability, which is a rapidly advancing technology and a cornerstone of Intelligent Transportation Systems (ITS). CAVs are equipped with Electronic Control Units (ECUs) dedicated to communication. These ECUs facilitate communication within the vehicle's onboard network and with external entities through the outboard network. The high level of the CAV system and scenarios that can be adapted to each use case and each CAV architecture (See more details in the paper [3]). The stakeholders in this system are (see Tab.3.): Users, Access Points (e.g. OBD II port, telematic control unit (TCU), multimedia, TPMS (tire pressure monitoring system), etc.), HPC, and Entities (e.g. service provider (SP), other vehicles and Road Side Units (RSUs)). Fig. 2 shows a general architecture of the data flow Onboard/Outboard CAV network, which is also the data flow diagram (DFD).

In the Onboard vehicle network, communication relies on various types of bus systems such as LIN, CAN, Ethernet, MOST, FlexRay, and others [6]. These bus systems enable internal communication among the different components within the vehicle. On the other hand, the Outboard vehicle network utilizes different technologies for transferring sensitive data. These include Dedicated Short Range Communications (DSRC), Wireless Access in Vehicular Environments (WAVE), WiFi, cellular networks (such as LTE, 4G, and 5G), Zigbee, Bluetooth, WiMAX, Ultra WideBand (UWB), and Radio Frequency Identification (RFID) [7]. Overall, the communication capabilities of CAVs encompass both the Onboard and Outboard networks, enabling seamless connectivity and data transfer for various applications and services.

In addition, the circulation of data through the different communication interfaces mentioned in the previous paragraph is flexible depending on the service needed. According to the state of the art [8, 9], the applications can be classified in three types: First, *IN-IN category*, which means that the data collected by the vehicle remains, is analyzed, processed, and stored in the vehicle. Second, *IN-OUT category* which includes use cases where the data collected in the vehicles are communicated to external entities. Third, *IN-OUT-IN category* are applications that send data to an external entity for processing, analysis, and collection. Then waits to receive the results of these operations. We recommend that readers refer to Section 2 of the article [3] for more detailed information.

In line with the approach taken in [4], we have identified the critical elements that can be exploited by malicious entities,

Table 1. DFD elements in the CAVs

DFD Elements Types	DFD Elements Members
Access points	OBD port, multimedia(USB), TCU, TPMS, Electronic vehicles charging (EVC), etc.
Users	Passenger, Owner
Data flow	AP-AP and AP-HPC through Onboard CAV network (i.e. CAN, ethernet, flex-ray, LIN, etc.), HPC - Entities and HPC - Users and Entities - Users through outboard CAV Network (cellular network, WiFi, multimedia **(Mult)** (Bluetooth), GSM network, etc.)
Datastore	Entities datastore (i.e. service provider DB, services DB, mobile phone DB, etc.) HPC datastore (also called CAV datastore)
Process	Entities (i.e. service provider **(Prov)**, services **(Ser)**, mobile phone **(Mobl)**, etc.), HPC

Table 2. Mapping privacy threats to DFD elements members(TP:Threat privacy, Enti: entities, U: Users)

Threat privacy	U	Enti	Enti DB (DS)	Bus Syst (DF)	Cellular (DF)	GSM (DF)	WiFi (DF)	OBD (AP)	Mult (AP)	TCU (AP)	GPS (AP)	TPMS (AP)
Linkability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Identifiability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓
Non-repudiation	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		
Detectability	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	
Disclosure of information	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		

thereby posing a privacy risk (see the paper [3] for more details). In the initial step (Step 1), we selected the relevant Data Flow Diagram (DFD) elements from the overall CAV architecture (refer to Table 3.). Subsequently, we established a mapping between LINDDUN privacy threats and the identified DFD elements (as shown in Table 3.). Each DFD element is associated with one or more privacy threats, which are determined by the type of DFD element. It is important to note that we have examined each 'x-mark' in the table during the next step.

In the following subsection, we will process the LINDDUN GO threat cards to termine the misuse case scenarios (**Step 3**). These threats will be classified and modeled according to the LINDDUN methodology. In this work, we focus solely on hard privacy concerns since they are the elements over which we have control through the mechanisms employed in the use case. Soft privacy, on the other hand, depends on the company's discretion and choices. In light of the key points outlined in section 2.2., we'll analyze each privacy threat associated with CAV, in the following subsection.

3.1. Linkability threat

*Linkability (**L**)* is the possibility that an adversary can associate two or more pieces of information about a user (e.g. messages, actions). LINDDUN GO [5] enumerates five linkability threats that encompass the Linkability of user data (L1), Linkability through distinguishable patterns (L2), Linkability of user requests (L3), Profiling users (L4), and Linkability of dataset (L5). By following the significant key points described in Section 2.2., we provide the following threat analysis.

Let's consider the hotspot as the inbound user with personal data, which refers to the moments when users register or connect to the CAV services using their IDs. The threat of the Linkability of user data (**L1**) can occur when the users' data are circulated inboard or outboard of the CAVs as described in the table 3.. As we know, when individuals register for CAV services or create user accounts, they typically provide personal information such as their name, contact details, and potentially even sensitive data like identification numbers or driver's license information. This user account information serves as a link to the users' data. For instance, External entities, such as server providers or computing servers may have an interest in attempting to link received data to user identifiers. These data are sent for specific purposes to obtain certain services. Or, a malicious third party may intercept the communication remotely or physically. The L1 can be applicable by employing specific attacks described in our previous papers [3]. The attacks can target the onboard CAV network or exploit vulnerabilities in V2X communication. The important scenarios are described in table 3 of the paper [3]. If user patterns are not encrypted or anonymized, server providers have the ability to link these patterns to users' identifiers. However, this threat increases the potential for other entities to obtain sensitive information about the users (see section 2.1.).

In CAVs, user data can potentially be linked by distinguishing patterns from data collected/ received from others, especially in the context of using the services proposed by companies (i.e. Pay-As-You-Go insurance[1], Platooning [10]). The threat is named Linkability through distinguishable patterns (**L2**). Considering the same hotspot and threat source as L1, there are a few scenarios where user patterns can be distinguished. The same as L1, each vehicle participating may have unique identifiers. These identifiers can differentiate and link data from specific vehicles to their users. For instance, the data collected from GPS

[1] https://www.theaa.com/car-insurance/advice/pay-as-you-go-car-insurance

or other positioning systems can reveal the location and movement patterns of vehicles within the use case. By analyzing these patterns and collecting them together, it may link specific user data to individual vehicles. Additional contextual information, such as the timing and location of vehicle entry or exit from the service, can further aid in linking user data. By analyzing this information and other data points, it may be possible to discern distinctive features specific to individual users.

The user requests can be linked by combining multiple attributes. This threat type is named Linkability of user requests (**L3**). For this threat type, we consider the same hotspot and threat source as L1. The applicability can be grouped when users interact with the service provider or make specific requests. The content of the user's request itself can provide distinctive attributes. Specific details, preferences, or requirements expressed in requests can be analyzed to identify patterns and link requests from the same user. Multiple request attacks are described in [11].

Profiling users threat (**L4**) is when the users can be profiled by analyzing their data for patterns. There are patterns derivable from the data. We consider the same hotspot and threat source as the threat L1. The threat is based on analyzing the collected data, patterns related to driving behavior, traffic conditions, vehicle performance, or user preferences may emerge. For example, the system might identify patterns such as certain driving styles during specific times of the day and speed variations. Or, by applying pattern recognition machine learning algorithms to the user data, certain characteristics or behaviors can be identified. For example, the system may discover that a particular user tends to maintain consistent speeds, prefers certain routes, or sleeps during autonomous driving. The articles [3, 12] describe multiple vulnerabilities and attacks.

The users are under the threat of having their stored information data linked to themselves. This means that there is a risk of identifying and connecting the stored data to specific individuals, and this threat is referred to as the Linkability of the dataset (**L5**). The hotspot and threat source is the same as the threat L1. In the context of CAVs use cases, there are specific areas where this threat can occur, such as the CAV dataset and the entities dataset 3.. Regarding the CAV dataset, one potential scenario is where a curious passenger or a malicious entity gains physical access to the datastore of the vehicles. This unauthorized access gives access to all the data collected/ received/ stored in the vehicle. Since the attacker has the whole data in plaintext (i.e. not encrypted). It can apply multiple machine learning attacks described in the paper [13]. Similarly, in the case of entities' datasets, the threat of linkability arises depending on the type of entity involved. The first type is server providers or server computing entities may have access to the data and could be curious to gather information about their clients. Since they have the user account information, They can link between the stored data and the individual. In addition, CAV systems may assign unique identifiers or vehicle identification numbers (VINs) to each vehicle participating in the use case. If these identifiers are associated with personal data during the registration or configuration process, the stored data can be linked to the individuals.

The second type is Devices connected to the CAV containing personal data stolen, such as mobile phones. The malicious entity gaining access to the datastore can further contribute to the linkability of the stored data (see the papers [13]). It is important to note that the potential entities involved in the linkability threat may vary depending on the specific use case considered. Therefore, assessing and addressing the privacy risks associated with each unique scenario is crucial. In the paper [14], the authors present various threats and attacks abusing the sensors of smart devices for malicious purposes.

3.2. Identifiability Threat

*Identifiability(**I**)* is the fact that an adversary can sufficiently identify the subject associated with an item of interest. LINDDUN GO [5] enumerates five Identifiability threats that are kind of the same as linkability threats. These threats encompass the Identifiability of user data (I1), Identifiability in user requests (I2), Identifiability in data (I3), Identifiability in data requests (L4), and Identifiability in dataset (L5). Our following analysis is based on the same structure outlined in subsection 2.2.

Most of the identifiability threats are similar to Linkability threats. Specifically, threats I1, I2, and I5 share the same hotspot, threat source, and applicability as threats L1, L3, and L5, respectively. Identifiability in data (**I3**) means the possibility of sufficiently revealing the user's identifier through all the data sent to the system. This means that the information transmitted by the user can potentially disclose their identity or unique identifier (e.g. the hotspot is the inbound personal data). The threat I3 depends principally on the organization of the systems itself. L3 is an applicable threat. Because the inbound dataset may contain user-specific information, such as personal details, vehicle identification numbers, or unique identifiers. By analyzing this data, it becomes possible to identify the dataset to the user.

In CAV services, identifiers in data requests can play a significant role in revealing the identity of an individual. The presence of unique identifiers used in interactions or when referring to data of an individual increases the risk of identification (e.g. Identifiability in data requests (**L4**)). The threat occurs in the inbound users with personal data, and the threat source is the same as L1. This threat is applicable in the sense that the vehicle can be identified if it communicates with the servers with the same ID, where the data are not encrypted, such as in Vehicle to Everything (V2X) communication. In the article [15], the author presents a three-layer framework (sensing, communication, and control) through which automotive security threats can be better understood.

3.3. Non-Repudiation Threat

*Non-repudiation(**NR**)* means that a user or an entity (i.e. services provider (SP)) cannot deny their involvement or the authenticity of a communication or transaction. LINDDUN GO specifies five Non-repudiation threats: Non-repudiation of service usage (Nr1), Non-repudiation of sending (Nr2), Non-repudiation of receipt (Nr3), Non-repudiation of storage (Nr3) and Non-repudiation of hiding data or metadata.

NR1 requires the detection of identifiers. Users cannot deny having used a service due to the presence of authentication and access logs. This is because users must authenticate themselves before accessing the service, and their access and usage activities are logged and recorded. The authentication process establishes the user's identity, meaning their actions within the use case can be traced back to them. Additionally, the logging of access and usage activities creates a record of the user's interactions with the service, making it difficult for them to deny their usage.

Otherwise, malicious users can take advantage of the offered services without paying anything.

In the CAV use case, users must not be able to deny their message sending (Non-repudiation of sending **Nr2**). The source of this threat can be attributed to the misuse in the use of cryptographic techniques, which are part of the system design and implementation. When users send a message within the use case system, there must be measures in place to establish the origin and authenticity of the message. One possible threat is the compromise of sender credentials. If an attacker gains unauthorized access to these credentials, he can impersonate the legitimate sender and falsely sign messages, leading to fraudulent activities or denial-of-service attacks. Another threat is the manipulation or alteration of the message during transmission. If an attacker intercepts the message and alters its content without being detected, he or she can manipulate evidence of the sender's intent and create disputes about the message's authenticity or integrity (see Table 3 [3]).

On the other hand, the entities (i.e. SP or service computing (SC)) must not be able to deny the receipt of the message (Non-repudiation of receipt **Nr3**). The threat source is similar to the threat Nr2. The possibility of a user denying or disputing the receipt of a particular Information of users. Malicious entities can deny receiving sensitive data in order to evade responsibility for it, which could lead to disputes or legal issues.

In addition, the entities are unable to deny or dispute the validity or existence of certain claims or statements regarding users' data that is stored during a use case (Non-repudiation of storage **Nr4**). The users' data's integrity and the actions related to it are securely stored and can be reliably traced back to the SP. The threat source depends on the manner of storing the user's data. Without any non-repudiation mechanism in place, malicious entities may be able to engage in various unethical or harmful activities with stored user data. For instance, The malicious SP may sell/ alter or manipulate the stored user data to suit their purposes. They can modify data or introduce false information, which can lead to potential harm or misinformation.

Finally, during the processing of the data, the entities must not be able to deny the processing of data (Non-repudiation of processing **Nr5**). A malicious entity can process the user's data beyond what is necessary or requested, resulting in excessive processing or unauthorized use of the data. This can lead to privacy violations and potential misuse of sensitive information.

3.4. Detectability Threat

*Detectability **(D)*** is that the adversary can sufficiently distinguish whether an item of interest exists or not. LINDDUN GO list four types of threat, which are :

First, the Detectability of users (**D1**) in the context of CAV services refers to the potential for users to be detected through the outbound flows of the system. Unlike the previous threat, the source of D1 does not depend on the techniques employed within the system but rather on external attack methods. Specifically, this threat is associated with passive attacks, where an adversary secretly intercepts or monitors data without actively modifying or disrupting the communication. For a more comprehensive understanding of passive attacks, I highly recommend reading the survey article [16]. Additionally, another malicious user with access to network traffic data may correlate outbound flows with other data sources to identify the clients. The malicious user can actively participate in the offered service.

Second, the detectability of service usage (**D2**) refers to the ability of an adversary to observe or detect when users are accessing or utilizing CAV services. This threat typically arises from external methods, similar to D1. It is applicable when an adversary can monitor network traffic, analyze data packets, or intercept communication channels to detect the communication between a service and its users. By observing patterns and timing, the adversary can infer the usage of the services offered. This threat leads to gaining insights into individuals' personal lives, daily routines, interests, or private interactions.

Third, the Detectability of an event (**D3**) refers to how an adversary can identify specific events or actions performed within the CAV application or system. The threat of detecting application events in CAV services can occur in both the inbound and outbound data flows. In CAVs services, there are several decisions taken by the entities that provide the service, such as the coordination or computation of intersection itineraries of multiple vehicles through servers. The privacy of these decisions is of paramount importance to ensure the security and confidentiality of the system. Adversaries should not be able to observe or interfere with these decision-making processes or access the sensitive information exchanged during these interactions. Only the necessary information required for making the decisions should be shared among the users.

Fourth, Detectability of records (**D4**) means that an adversary may detect the existence of specific records within the system. Generally, CAVs services store various records, such as user profiles, trip histories, or vehicle performance data. If an adversary can detect the presence or absence of certain records, it could provide them with insights into the presence of certain data or records, which may reveal sensitive information or provide indications about the activities or behaviors of the users.

3.5. Disclosure of Information Threat

The Disclosure of Information threat refers to the unauthorized exposure or dissemination of personal information to individuals who are not authorized to access it. In the context of privacy threats, LINDDUN GO identifies five specific threat types that need to be explored:

Firstly, we have the Excessively sensitive data disclosed (**DD1**), this threat focuses on the possibility of processing more sensitive or fine-grained data than necessary for the functionality of the use case proposed. The threat occurs when data should be sent to the entities that offer the CAvs service. Usually, this issue arises from the mechanism used in the system where there is a lack of verification or validation of the type of data that is needed or necessary. The malicious service provider can request users to provide multiple pieces of data, knowing that only a portion of it is necessary for the required service.

Secondly, we delve into the Excessive (**DD2**) amount of data disclosed, which focuses on the situation where the system processes and discloses a larger quantity of data than what is actually required for its intended functionality. The hotspot and the threat source are the same as the threat DD1. A malicious SP can unnecessarily handle and expose more data than necessary, which can pose privacy and security risks.

Moving on to the third threat, Unnecessary Processing (**DD3**) refers to situations where data undergoes additional processing, analysis, or enrichment that is not necessary for its intended functionality. This means that the data is subjected to unnecessary

Fig. 3. Mapping of the privacy requirements to the Privacy Enhancing Solutions (PES).

PETs			Elementary privacy objectives														
			I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV	XV
Processing Layer	Secure and outsourced	Homomorphic Encryption	✓	✓		✓	✓		✓	✓					✓	✓	
		Secure Multiparty Computation															
		Zero-knowledge proofs															
	Anonymization	Differential privacy	✓	✓		✓	✓		✓	✓		✓	✓				
		Syntactic															
		Perturbation															
Communication Layer (Data Flow)		Symmetric and Asymetric encrypt	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓
	Anonymization	Mixnet Protocol	✓		✓	✓		✓	✓		✓	✓		✓			✓
		Tor-related Protocols															
		Random Walk/DHT Protocols															
		DCnet Protocols															
	Authentication	Digital Signature	✓		✓	✓		✓	✓		✓			✓	✓		✓
		Certificateless signature															
		Quantum-safe Signatures															
		Homomorphic Signature															

operations or transformations that go beyond what is essential or relevant for its intended purpose. The threat occurs during the processing stage of data. Similarly to DD1, this threat can originate from the mechanisms or techniques employed in the system. From the perspective of the entities involved, such as service providers (SP) or service consumers (SC), the processing of data may involve unnecessary operations or steps that exceed what is required for the intended functionality, potentially introducing privacy risks or unauthorized access to sensitive information.

Next, we consider the threat of Unnecessary Storage (**DD4**), which centers around the situation where data is stored for a longer duration than necessary. This means that data is retained or kept in storage beyond the required timeframe for its intended purpose. In the context of CAV use cases, the threat of DD4 can manifest in the CAV datastore or entities datastore. Similar to the threat DD1, the main source of this threat lies in how the data is stored. It raises questions about the mechanisms employed for data storage, such as whether they are encrypted or anonymized. Improper management or disposal of a data store after it has fulfilled its intended purpose can give rise to potential privacy risks and heightened exposure to sensitive information. It is crucial to ensure that data storage practices align with data retention policies and follow best practices to protect privacy.

Lastly, we explore the threat of Overbroad Exposure of Personal Data (**DD5**), which investigates the situation where personal data is shared with more services or external parties than necessary for the intended functionality or purpose. DD5 occurs when personal data is moved outside the CAV use case. The source of this threat may come from the weakness of the mechanism used or an external factor that accesses the data and shares it with multiple servers. This threat typically arises from weaknesses in the mechanisms used for data handling or external factors that gain access to the data and share it with unauthorized services. A malicious entity may illegally sell the database of users to other companies without the consent of the owners. Additionally, a malicious third party may gain unauthorized access to sensitive data by exploiting one of the attacks presented in the papers [3, 14].

After presenting the relevant threat scenarios for data flow in CAV, in the paper [3], we present the high-level privacy objectives that map the threat scenarios to specific privacy requirements. They are: Unlinkability of Users Data Flow (**I**), Unlinkability of users during Processing (**II**), Unlinkability of users through Access Point (**III**), Anonymity and pseudonymity of users Data Flow (**IV**), Anonymity and pseudonymity of users during Processing (**V**), Anonymity and pseudonymity of users through Access Point (**VI**), Plausible deniability of users Data Flow (**VII**), Plausible deniability of users during Processing (**VIII**), Plausible deniability of users through Access Point **IX**, Undetectability of users Data Flow (**X**), Undetectability of users data during Processing (**XI**), Undetectability of users data through Access Point(**XII**), Confidentiality of users Data Flow (**XIII**), Confidentiality of users during Processing (**XIV**), and Confidentiality of users through (**XV**). For more detailed information, we refer the reader to our previous papers [3]. As for the final step of the LINDDUN methodology, Table. 3 showcases the Privacy Enhancing Technologies (PETs). These PETs are derived from a systematic literature review proposed in [cite: PET], and they are mapped to the identified privacy objectives.

4. Discussion and Conclusion

In this section, we consider the implications of our privacy analysis within the broader context of research on privacy-respecting connected and autonomous vehicle systems. The data we have gathered provide valuable insights for privacy threat analysis and give rise to multiple interpretations and significant considerations. This work builds upon and enhances our previous research by introducing several new elements. In our previous work [3], we initiated a modeling process for the CAV system to address threats to the flow of personal data within the Onboard/Outboard networks. This process involved categorizing the data flow based on the considered use case (IN-IN/IN-OUT/IN-OUT-IN categorizations), classifying CAV use cases, identifying potential attack vectors in general, and assessing some privacy threats in CAVs. To extend upon our previous research, the present work defines the CAV data that needs protection and presents a classification of this data. Furthermore, this work provides a comprehensive elucidation of privacy threats in CAV use cases, utilizing threat-type cards. Unlike our prior work, which concentrated solely on threats linked to data flow likability, the current paper encompasses all threats associated with CAVs based on the various recent attacks identified in our previous research. Table 3. shows the read-make analysis conducted in

this work, in comparison to the previous table in [3]. The current findings of this study enhance our capacity to proactively anticipate potential attacks before they manifest. Through this analysis, our primary objective was to identify potential attacks in advance, with a specific emphasis on data-driven privacy, thereby ensuring compliance with GDPR requirements.

One of the main distinctive point of the present paper from the related works is its field of application. As detailed in [17], the authors introduce a systematic approach aimed at enhancing the comprehensiveness of threat modeling. This systematic method is exemplified through the utilization of the well-established LINDDUN threat modeling methodology, which is applied to four pivotal pieces of literature concerning privacy threat modeling in the automotive sector, including our type of work. To the best of our knowledge, no attempt has been made to analyze the privacy threats in the CAVs field that follows a privacy threat analysis methodology.

Our effort lays a robust groundwork for the broader task of tailoring privacy risk assessments specifically for businesses operating in the automotive industry. These companies are compelled to adhere to GDPR principles [1]. By aligning their work with this initiative, they can effectively address all GDPR principles, making their solutions market-ready.

It is important to acknowledge that the methodology is time-consuming and requires expertise. This is primarily due to the evolving nature of misuse threats, which may present more complex scenarios for each specific use case. Additionally, care must be taken to avoid overlooking any potential threats to privacy during the creation of threat cards. In our future work, we plan to focus on the risk-based prioritization aspect of this methodology (Step 4). This entails developing a risk rating methodology for calculating, evolving, and assessing privacy risks associated with the CAV misuse case discussed above.

Acknowledgements

This research is funded by the German-French joint collaboration program on “Cybersecurity” funded by the MESRI (via ANR) and BMBF (via VDI/VDE IT) (ANR-20-CYAL-0008-01).

References

[1] General Data Protection Regulation. General data protection regulation (gdpr). *Intersoft Consulting, Accessed in October*, 24(1), 2018.

[2] Yod-Samuel Martin and Antonio Kung. Methods and tools for gdpr compliance through privacy and data protection engineering. In *2018 IEEE European symposium on security and privacy workshops (EuroS&PW)*, pages 108–111. IEEE, 2018.

[3] Badreddine Chah, Alexandre Lombard, Anis Bkakria, Reda Yaich, Abdeljalil Abbas-Turki, and Stéphane Galland. Privacy threat analysis for connected and autonomous vehicles. *Procedia Computer Science*, 210:36–44, 2022.

[4] 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.

[5] Kim Wuyts, Laurens Sion, and Wouter Joosen. Linddun go: A lightweight approach to privacy threat modeling. In *2020 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*, pages 302–309. IEEE, 2020.

[6] Miroslaw Staron, Miroslaw Staron, and Darko Durisic. Autosar standard. *Automotive Software Architectures: An Introduction*, pages 81–116, 2017.

[7] Xiaoqiang Sun, F Richard Yu, and Peng Zhang. A survey on cyber-security of connected and autonomous vehicles (cavs). *IEEE Transactions on Intelligent Transportation Systems*, 23(7):6240–6259, 2021.

[8] Cheng Huang, Rongxing Lu, and Kim-Kwang Raymond Choo. Vehicular fog computing: architecture, use case, and security and forensic challenges. *IEEE Communications Magazine*, 55(11):105–111, 2017.

[9] M Muzakkir Hussain, Mohammad Saad Alam, and MM Sufyan Beg. Fog computing model for evolving smart transportation applications. *Fog and Edge Computing: Principles and Paradigms*, 22(4):347–372, 2019.

[10] Badreddine Chah, Alexandre Lombard, Anis Bkakria, Abdeljalil Abbas-Turki, and Reda Yaich. H3pc: Enhanced security and privacy-preserving platoon construction based on fully homomorphic encryption. In *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023.

[11] Nazrul Hoque, Monowar H Bhuyan, Ram Charan Baishya, Dhruba K Bhattacharyya, and Jugal K Kalita. Network attacks: Taxonomy, tools and systems. *Journal of Network and Computer Applications*, 40:307–324, 2014.

[12] Mani Amoozadeh, Arun Raghuramu, Chen-Nee Chuah, Dipak Ghosal, H Michael Zhang, Jeff Rowe, and Karl Levitt. Security vulnerabilities of connected vehicle streams and their impact on cooperative driving. *IEEE Communications Magazine*, 53(6):126–132, 2015.

[13] Maria Rigaki and Sebastian Garcia. A survey of privacy attacks in machine learning. *ACM Computing Surveys*, 2020.

[14] Amit Kumar Sikder, Giuseppe Petracca, Hidayet Aksu, Trent Jaeger, and A Selcuk Uluagac. A survey on sensor-based threats and attacks to smart devices and applications. *IEEE Communications Surveys & Tutorials*, 23(2):1125–1159, 2021.

[15] Zeinab El-Rewini, Karthikeyan Sadatsharan, Daisy Flora Selvaraj, Siby Jose Plathottam, and Prakash Ranganathan. Cybersecurity challenges in vehicular communications. *Vehicular Communications*, 23:100214, 2020.

[16] Qiyi He, Xiaolin Meng, and Rong Qu. Survey on cyber security of cav. In *2017 Forum on cooperative positioning and service (CPGPS)*, pages 351–354. IEEE, 2017.

[17] Mario Raciti and Giampaolo Bella. How to model privacy threats in the automotive domain. In *Proceedings of the 9th International Conference on Vehicle Technology and Intelligent Transport Systems (VEHITS 2023)*, 2023.