Emerging Privacy Challenges and Approaches in CAV Systems

Ugur I Atmaca[1*], Carsten Maple[2], Mehrdad Dianati[3]

[1]Doctoral Researcher, WMG, University of Warwick, Coventry, UK
[2]Professor of Cyber Systems Engineering, WMG, University of Warwick, Coventry, UK
[3]Professor of Autonomous and Connected Vehicles, WMG, University of Warwick, Coventry, UK
**ugur-ilker.atmaca@warwick.ac.uk cm@warwick.ac.uk, m.dianati@warwick.ac.uk,*

Keywords: Connected and autonomous vehicles, intelligent transportation systems, privacy.

Abstract

The growth of Internet-connected devices, Internet-enabled services and Internet of Things systems continues at a rapid pace, and their application to transport systems is heralded as game-changing. Numerous developing CAV (Connected and Autonomous Vehicle) functions, such as traffic planning, optimisation, management, safety-critical and cooperative autonomous driving applications, rely on data from various sources. The efficacy of these functions is highly dependent on the dimensionality, amount and accuracy of the data being shared. It holds, in general, that the greater the amount of data available, the greater the efficacy of the function. However, much of this data is privacy-sensitive, including personal, commercial and research data. Location data and its correlation with identity and temporal data can help infer other personal information, such as home/work locations, age, job, behavioural features, habits, social relationships. This work categorises the emerging privacy challenges and solutions for CAV systems and identifies the knowledge gap for future research, which will minimise and mitigate privacy concerns without hampering the efficacy of the functions.

1 Introduction

The Internet of Things (IoT) is promising revolutionary changes in the way people live, work, transport, and interact with technology by bringing together multiple sensors, actuators, communications technologies, data and processing. One of the significant areas of IoT development is in the area of intelligent transportation systems (ITS) [1]. ITSs consist of a network of roadside units (RSU), vehicular on-board electronic control units (ECU), distributed computing and storage systems [2]. Wireless networks communications such as V2V (Vehicle-to-Vehicle), and V2I (Vehicle-to-Infrastructure) are enabled through technologies such as IEEE 802.11p DSRC/WAVE (Dedicated Short Range Communication/Wireless Access in Vehicular Environments) and cellular advances such as C-V2X [3].

Modern vehicles are evolving to be safer, more energy efficient, more comfortable and accessible by being equipped with a wide range of sensors and ECUs. Developments in wireless communication, sensing the internal and external surroundings, and capability of decision taking for driving are advancing the state of the art in connected and autonomous vehicles (CAVs). To help categorise the level of autonomy in CAVs, the Society of Automotive Engineers (SAE) has created a set of standardised levels for autonomy from level 0 to 5, wherein at level 5 the vehicle is expected to take all driving decisions without any user monitoring required [4]. Further to the autonomy advances, the development of reliable low latency wireless communications (i.e. the envisioned 5G), and cloud-based infrastructure are able to coordinate and increase the knowledge-base of CAVs. Cloud-assisted CAVs bring advantages such as broader connectivity for real-time traffic optimisation realised through the cloud and mobile edge computing [5].

As well as promising numerous benefits, CAVs also the potential for negative consequences such as privacy invasions and tracking [6]. While we mention privacy invasions, it is important to recognise that there is not a unified definition of privacy. Indeed, there have been propositions that there is a paradox [7], and that privacy varies according to different people, communities, and cultures [8]. From a business perspective, privacy compromises might negatively influence the reputation of manufacturers and businesses. Therefore, privacy protection is an essential aspect of consumer trust, and hence the adoption of new technology [9], [10]. Approaches to protect privacy can be termed Privacy-enhancing technologies (PETs) and they can promise significant advances in data sharing to drive all kinds of applications forward with confidence.

Privacy challenges in CAV systems are can be analysed considering identity anonymisation and authentication in vehicular ad-hoc networks (VANETs) by using schemes such as pseudonym-based privacy and group signatures. Recently, Qu et al. addressed and classified privacy challenges and requirements for vehicular networks [11]. However, privacy challenges that might arise out of the CAV functions (e.g. location-based services, etc.) have not been included. If there is a known link between two points of data, anything learned by one of the points might enable to make an inference about the other data point. Asuquo et al. analysed the privacy requirements, challenges and approaches Location-Based Services (LBS) in vehicular networks [12]. Ni et al. analysed the privacy requirements of fog-computing based vehicular systems and their functions [13]. To the best of our knowledge, this work is the first review, which delivers a

comprehensive analysis of the privacy challenges of both vehicular networks and CAV functions. It proposes a novel taxonomy and discussed the identified state-of-the-art in Differential Privacy (a particular type of PET that is attracting significant attention) regarding its applicability to the privacy challenges in CAV functions we have identified. This work also identifies the existing knowledge gap in the research base and proposes emerging research directions.

2 Methodology

In this work, we provide a brief background for existing notions of privacy and characteristics of CAVs. We then present the privacy challenges and privacy-preserving techniques, which have been identified through a semi-systematic review and qualitative analysis. We consider research published between 1998, the year Latanya Sweeney's seminal work on k-anonymity [14] was published, and 2018. Finally, we discussed the potential applicability of the state-of-the-art privacy-preserving approaches for the privacy challenges and in CAV systems.

2.1 Paper Review Protocol

The following table summarises the paper review protocol employed in this study, aimed at reducing the potential risk of selection bias.

Table 1 Paper review protocol

Criteria	Description
Timeframe	Between 1998 and 2018
Language	English
Aims	• To identify privacy challenges in ITS and CAV functions. • To establish the state-of-the-art PETs and evaluate their potential on the application domain.
Topic	"Privacy Preserving Techniques", "Vehicular Communication", "Intelligent Transportation Systems", "Connected and Autonomous Vehicles"
Type	Academic Journals, Peer Reviewed Conference Proceedings
Search Technique	Boolean and word combinations
Keywords	Privacy preserving, Intelligent Transportation Systems, Connected and Autonomous Vehicles, Differential Privacy.
Databases	IEEE Xplore, Science Direct, ACM Digital Library, Google Scholar.
Include	The highly cited papers and the recent 3 years` papers are prioritised.

3 Notions of Privacy

The concept of privacy is not new, but it is elusive. In the early stages, privacy was firmly aligned with secrecy, as the Code of Hammurabi has brought the protection of the house of every Ancient Babylonian against others` intrusion. In the 1800s, privacy was defined as the "right to be let alone" [15]. The definition has been expanded and developed as human`s needs changed over time. It has been defined as an "umbrella term" which refers to the group of wide and distinct elements [16], containing "bodily, communications, territorial, and informational" elements under the umbrella. Bodily privacy is the physical protection of individuals and is profoundly linked to safety. Privacy of communications represents the confidentially of the information transmitted via any communication channel. Territorial privacy is controlling the intrusion into personal territories such as the home, workplace, vehicle and public spaces. Informational privacy indicates personal data aggregated by organisations and companies [17]. Today, the International Covenant on Civil and Political Rights currently preserves it as one of the legal and human rights in many nations since 1966.

Privacy concerns change with the rapid advancement of technology such as the increased opportunity for data collection, storage and computation. Dwork noted that privacy could be disclosed when a link is established between different data points of the same data owners [18]. The ideal privacy for the information systems is defined as "nothing about an individual should be learned by an adversary from the database that cannot be learned without access to the database" [19]. Although it is not entirely achievable, it stands as a utopic goal of privacy mechanisms.

The recent EU General Data Protection Regulation (GDPR) has asserted that privacy should be taken into account from the early stage of system development rather than implementing privacy protection mechanisms later. The term "Privacy by Design", has been coined to describe this principle [20]. One recent survey state that the understanding of privacy is changing for different people, communities and cultures, which that this should considered when developing privacy mechanisms [21].

Privacy, as a notion, has been evolving differently from the normative and technical points of view. Westin defined privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others." [22]. Our privacy consideration is based on this definition. CAV users need to share the data about themselves to be able to make maximum use of advanced cooperative autonomous driving functions. In general, increasing the privacy of data comes, in general, as a cost in terms of sacrificing some of the advantages accruing from the functions. However, it is an open research direction as to what the optimal balance of privacy versus function efficacy should be.

4 Characteristics of CAVs

VANETs have highly dynamic network structures, comprising wireless communication technologies between RSUs installed along roads, and the On-Board Units (OBUs) installed in vehicles [2]. The communication can be

established either V2V and V2I [23]. VANETs have evolved to into ITSs through the increased connectivity and reliable infrastructure services, which can be also supported by cloud services [24]. ITSs have dynamic, heterogeneous, distributed and open nature, leading to have distinctive features and specific requirements. The ITS features and security requirements are represented on the following Table 2 and Table 3 (adapted from [25]).

Vehicles are evolving to become CAVs through being equipped with a wide range of sensors (e.g. Internal and External Cameras, Lidar, Radar, Ultrasound Sensors, GPS, etc.) and hundreds of ECUs [26]. CAVs generate a massive amount of rich-dimensionality data, which is utilised by a variety of functions and shared with other applications. This results in a need for ensuring privacy protection for CAV users, regarding data, identity and location privacy. It should be noted that these functions may not only contain personal sensitive data but also commercial and research sensitive data.

Table 2 ITS Features [25]

Features	Description
Powerful capacity	It includes powerful units in terms of energy resource, localisation, computation, storage and data rate capabilities.
High mobility	It includes many mobile units with different speed and direction.
Dynamic network topology	The units can join and/or leave the network very quickly depending on their location and speed.
Time sensitivity	The latency is one of the most important limitations (e.g. 100ms delay for safety-related messages).
Sufficient energy	The unit has sufficient resource in terms of energy, storage and computation for implementation of complex algorithms.
Good physical protection	Each unit needs protection against physical attacks.
Unbounded network size	It is not bounded with a special area.
Wireless communications	The units have wireless comm. among each other.
Heterogeneous V2X communication	Vehicles use a wide range of communications technologies
Heterogeneous environments	Vehicles can operate in different environments.
Security and privacy	It needs security and privacy mechanisms considering the time constraint and the low computation complexity.

Table 3 ITS Security Requirements [25]

Requirements	Description
Authentication	It is being able to verify data sources and destinations
Data integrity	It is being able to verify and validate the received data is not maliciously altered.
Privacy and anonymity	The users and vehicles should not be able to identify or tracked by the exchanged messages.
Availability	The units should be available for real-time applications.
Traceability and revocation	The authorities should be able to detect malicious entities and verify their identities.
Authorisation	ITS should be able to control access for the messages or enabled functions.
Non-repudiation	ITS units should be able to associate with their messages and actions.
Robustness	ITS should provide robust safety against cyber-attacks.
Confidentiality	The transmitted data should be protected from malicious or unauthorised entities.

5 Privacy Challenges

As mentioned, CAVs create challenges in cybersecurity nd privacy, and there exist a number of research proposals for reducing the likelihood of cyber-attacks against CAVs. From the cyber security point of view, the importance of privacy protection is to reduce the potential information loss and reputation loss due to cyber-attacks. Users` privacy should also be protected from potential disclosures by service providers, governmental agencies and third-party entities. This section presents the privacy challenges of CAV systems. The privacy challenges both arising from participating in ITS and CAV functions are analysed through three subclasses: Data Privacy, Identity Privacy and Location Privacy.

In the following, we will discuss the privacy challenges present due to the inter-vehicular communication involved in ITS. The vehicles can be identified and tracked by messages propagated through the ITS. Usually messages are required to contain Vehicle ID or an appropriate pseudonym for authentication. Furthermore, it is standard that vehicles regularly broadcast safety awareness messages including location and direction. These messages can be eavesdropped to reveal vehicle identities or track vehicles. The privacy challenges in ITS are usually regarding identity and location but it is also required to ensure access of public authorities to identify and location information for accountability purposes [27].

ITSs involve different type of vehicles and different level of autonomies. The privacy challenges in ITS influences all of the vehicles, however, there are additional privacy challenges regarding the functions of CAVs, such as traffic planning,

optimisation, safety and cooperative autonomous applications (e.g. platooning). CAVs rely on data from various sources (e.g., on-board sensors, other vehicles, infrastructure, etc.) while much of this data is privacy-sensitive including personal (e.g. geo-locations, number plates, identity information, biometrics, etc.), commercial (e.g. trucks' loads, origin, destination information, etc.) and research (e.g. telemetry, test data, etc.). The share of this data might cause a privacy leakage of individuals or groups.

People might encounter many privacy challenges during the use of CAVs. However, some of these challenges are not specific to CAVs, but rather they are about being in a public place or in a taxi such as the privacy challenges due to vehicular external/internal cameras. In the following, we will discuss the perceived privacy challenges presents specifically due to the use of CAV functions. The open question here is that: what is the required dimensionality and accuracy of the data to provide sufficient efficacy of the functions, without significantly compromising privacy.

5.1 Route Planning

CAVs are able to offload the navigation routes to the cloud-based infrastructure in order to optimise the routes using real-time road information. This presents a location privacy challenge since the user needs to reveal information about the start and the end locations of travel and it is highly likely that one of these is the user`s home or workplace.

5.2 Participating in a Vehicle Platoon

The examples of cooperative movement exist in nature such as a group of migratory birds fly together in a sequential manner, and a group of dolphins swim together. The cooperative movement is beneficial for them regarding energy consumption, comfort and safety. Similar benefits exist for cooperative vehicles as well, such as using actual road infrastructure more efficiently, lowering energy demand by reducing aerodynamic drag for following vehicles, and improving comfort and safety [28]. A group of vehicles moving in a sequence with a minimal inter-vehicle distance is named a vehicle platoon.

In order to participate in a platoon, vehicles need to share several types of data (e.g. velocity, acceleration, destination, vehicle type and current location), depending on the optimisation objective of the platoon that the platoon management algorithms aims to maximise or minimise. From the literature review, it has been observed that most of the platoon management algorithms require receiving participating vehicles` destinations. The location is probably the most privacy-sensitive data. However, vehicle weight, required for the computation of maximal velocity and inter-vehicle distance, might be considered privacy-sensitive since it might reveal knowledge about the number of people in a car or the amount of load in a truck. Additionally, the correlation between origin-destination and truck weight has the potential to reveal commercially sensitive knowledge. Therefore, these applications present data privacy and location privacy challenges. However, the privacy challenges arising from participating in a platoon have not been widely discussed in the literature though designing platooning algorithms is receiving a great deal of attention. Amoozadeh et al. [29] discussed the privacy-preserving factor regarding the selection of platoon coordination strategy and decided to use the centralised approach since it would be fast, scalable and enhance privacy in some degree since the leader coordinates all communication, and only the platoon leader knows all configuration. The platoon configuration is only shared with the new leader when the old leader leaves. Clearly this approach is privacy-aware, though does not fully address all of the challenges.

5.3 HD/3D Map Updates

Many CAV functions are supported by a HD/3D Map, that provides information about the road and environmental conditions before they are detection by a vehicles on-board sensors. For instance, if an accident has happened on the road, the HD/3D Map can inform the vehicular navigation system to use alternative routes [30]. The integrity of the updates is vital. The updates will shape the map, and they should be credible and verifiable. However, the updates can also reveal the location and temporal data of the issuer CAV as well. This presents identity privacy and location privacy challenges.

5.4 Vehicular Telemetry and Biometric Data Collection

CAVs are equipped with sophisticated sensors to collect continuous data about the surrounding environment and the user. Analysing this sensory data can be used in fault detection [31], driving monitoring [32], and driver monitoring [33]. The biometric data of users can reveal information about personal health conditions [34]; likewise, telemetric data of trucks can reveal commercially sensitive information.

Vehicular telemetric data is highly valuable for many parties. From the perspective of users, it can help developing more safe, secure and comfortable vehicles and services. From the traffic authorities and governments` perspective, it can allow optimising the use of infrastructure. From the OEMs` perspective, it can be analysed to develop new products and services. According to McKinsey`s Monetizing Car Data report, the overall revenue of vehicle-generated data market might reach 450 - 750 billion USD by 2030 [35]. Although the data can be anonymised before using it, the de-anonymisation can be accomplished by cross-referencing [36]. Therefore, de-anonymised data might cause to identify, track and reveal some other private information of some users. The privacy challenges here can be associated with identity, location and data privacy.

6 Privacy-Preserving Approaches

Developing privacy-preserving approaches has gained significant attention in a variety of application domains. This section examines recent privacy-preserving approaches, which can be applied to CAV systems, and classifies them

into three categories: 1) Privacy Based on Anonymity, 2) Privacy Based on Perturbation, and 3) Privacy Based on Cryptography.

6.1 Privacy Based on Anonymity

In the generic structure, the data can be classified into attributes such as explicit-identifier, quasi-identifier, sensitive attributes, and non-sensitive attributes. The explicit-identifier set is information directly concerning identification, such as names, IDs or number plates. The quasi-identifier set is information that might potentially disclose the data owner. Sensitive attributes are private information related to the data owner, and non-sensitive attributes are all of the rest information including public information [37]. The vehicles are usually linked to their users or the owner companies (e.g. commercial goods carriage). Therefore, the users or the companies can be considered as the data owner of the vehicular data. It was originally thought that privacy preservation could be achieved by anonymising the data. However it has been proved, the exclusively anonymising the attributes is not a robust method of preserving privacy. The combination of non-sensitive attributes might disclose private information of the data owner [36]. Traditionally anonymisation is performed by a centralised and trusted system that does not attempt to identify the individuals or disclose sensitive information. However, it cannot be guaranteed which is generally defined as honest-but-curious cloud server in the literature.

One of the most well-known techniques is k-anonymity. It states that if a data entry in k data entries cannot be identified from the rest of the $k-1$ data entries, the dataset has k-anonymity [14]. In this method, an adversary can distinguish an individual with the maximum probability of $1/k$. The limitations of k-anonymity have been addressed, and the l-diversity method was developed as an extension [38] followed by t-closeness [39].

6.2 Privacy Based on Perturbation

Differential Privacy (DP) techniques are used to protect privacy by perturbing original data with a random noise while ensuring that the amount of distortion has little effect on the output. The primary purpose of DP is eliminating the change in the query outputs by the addition or removal of a single entry to the dataset. In the worst case, an adversary might have the background knowledge about all the data items except a single item that belongs to an individual. The query answers might help to disclose the data item. However, DP introduces uncertainty to the query answers. In this case, DP provides a privacy guarantee that the individual`s sensitive information will be not disclosed. In other words, it resolves the risk privacy disclosure by participating in a dataset for individuals [18]. According to Chen et al., DP is the state-of-the-art privacy notion, which provides provable privacy guarantee independent from an adversary's background knowledge and computational power [40].

Let D_1 be the dataset as being the collection of n entries from a universe S. Let the other dataset, D_2 to be a neighbouring dataset of D_1 (i.e. consists of same entries except differing only one entry), $\|D_1 - D_2\|_1 \leq 1$. The definition of DP is for the given neighbouring datasets D_1 and D_2 is that; a randomised mechanism M provides (ϵ,δ)-DP if the datasets $D_1$and D_1 are neighbouring datasets while all $S \in Range(M)$,

$$\Pr[M(D_1) \in S] \leq e^{\epsilon} \times \Pr[M(D_2) \in S] + \delta$$

The ratio of the two probability equations is bounded by e^{ϵ}. The ϵ is the privacy budget (denotation of the privacy loss) and the higher the value, implies providing less privacy guarantee. The probability of δ is the limitation of privacy violation and negligible in many settings. When the δ is equal to 1, the privacy guarantee is completely removed. When $\delta = 0$, the randomised mechanism ensures ϵ-DP which is the strictest privacy protection from the definition and generally called *pure DP*. On the other hand, When $\delta > 0$, it is called *approximate DP* [41].

The sensitivity is the parameter of a dataset for determining how much perturbation is required in the randomised mechanisms. If f is a query in dataset D_1 then the amount of perturbation in the output is calibrated with $f(D_1)$. Sensitivity is also depends on the type of the query f. The global sensitivity is used for *count* and *sum* queries, however, it returns high values for the queries like *median* and *average*. The local sensitivity is used for such queries.

DP has recently emerged as the gold standard of data privacy since it is not tied up with the background information of adversary and computation power. It does, however, introduce a trade-off between data utility and privacy. CAVs are highly mobile, resource constrained and time-bound and the functions should run accurately not to jeopardise the safety of users. For these reasons, implementing DP in the vehicular domain has challenges. We divided the use of DP into three classes, namely VANET, OEMs data, and CAV functions, and present a classification of current research in Table 4.

Table 4 Differential Privacy Studies in Vehicular Domain

Class	Ref.	Year	Privacy Mechanism	Outcome
VANET	[42]	2018	Machine learning based collaborative intrusion detection system	Used DP with dual variable perturbation. Investigated the security privacy trade-off.
	[43]	2018	DP for publishing mobility data for transportation applications.	Proposed Constraint-Based DP and evaluated its performance.

	[44]	2018	DP for publishing trajectory data.	Performance evaluation of the proposed mechanism.
OEM Data	[45]	2013	DP for ITS management to protect Floating Car Data	Used Laplacian noise with smooth sensitivity. Discussed the use of event-level DP.
	[46]	2017	DP mechanisms for vehicles	Discussed different mechanisms of DP.
CAV Functions	[30]	2018	Discussed privacy and integrity dependency for CAVs.	Discussed HD/3D Map Update scenario and DP among techniques.
	[47]	2018	Intelligent Route Planning	Proposed Scalable Privacy Mechanism.
	[48]	2018	DP for location data in Crowdsourcing applications.	Conducted experiments of the proposed algorithm.

6.3 Privacy Based on Cryptography

It has been discussed in the previous section that the information of vehicle identity and location are usually exchanged in V2V and V2I communications. Threat agents can analyse these messages to identify and track vehicles. In the literature, group signature based schemes can be used to provide unlinkability of the vehicular messages from the same owner for different events. These schemes provide conditional privacy that only the group manager can access the information of real vehicle identities [49]. Pseudonym change based schemes are another research area to provide unobservability for vehicles.

The challenges with using techniques based on cryptography are that the approaches generally introduce high computational and communicational overheads [50]. Thus there may arise a trade-off between privacy protection and safety, and this is considered be one of the open research areas [51].

Secure Multiparty Computation (SMC) and Homomorphic Encryption (HE) are also cryptographic techniques for privacy preserving. However, they have not been considered in the content of this study due to the following reasons: 1) Each party learns the result of some computation in SMC; there is no privacy guarantee that none of the parties is revealing any sensitive information received from the computation result. 2) HE enables computation directly on encrypted data;, but introduces high computational overhead [52].

7 Lessons Learned and Open Issues

The understanding of privacy changes for different people, communities and cultures, which should be considered when developing privacy mechanisms.

Earlier studies usually considered protecting identity privacy and location privacy in ITS by using pseudonym-based and group signature-based mechanisms. However, techniques based on anonymity fails under the cross-referencing and extra background knowledge situations [53].

Although there are limited studied of DP in real-world applications, DP has recently emerged as the gold standard of data privacy for two reasons [54]. First, DP introduces uncertainty for privacy protection is independent of background knowledge and computation power. Any further data analyses of differentially private results do not reveal information. This is the 'post-processing' feature of DP. Second, DP allows several queries to target the same data, known as the 'composability' feature [55]. One of the recent, local DP studies divided the users into two classes as 'opt-in' and 'opt-out'. By analysing the general tendency of a very small opt-in group (up to 10% of total users) in a centralised DP manner, the overall data utility remain higher than previous local DP studies [56]. Recent technological developments, such as predictive systems, automation systems, and artificial intelligence rely on data that is also applicable to CAV functions. By the nature of data, the perfect privacy protection can be only achieved by publishing no data, which is impossible for many CAV functions. The features of ITS, seen in Table 2, are the boundary conditions of CAV functions. Thus, the minimum data utility should be rigorously maintained in the privacy mechanism so as not to jeopardise the reliability of the service. A further challenge in privacy preservation is that privacy can be disclosed if a link can be established between different data entries of the same data owners. A potential solution for this is publishing data with an amount of randomisation (i.e. approximation) rather than publishing the exact data [18].

There are also knowledge gaps in DP that should be addressed. The 'privacy budget (ϵ)' should be set carefully to keep the required level of data utility. However, there is not a sufficient scientific foundation for this as it strongly depends on the dataset. Investigating data flows of the functions necessary to be able to assign the appropriate parameters, but more research is still needed regarding setting the parameters and their implications. DP for location data is an emerging research area. There are challenges due to distance-based sensitivity calculation [57], and sparsity of location dataset leads to adding a large magnitude of noise [58].

The identified privacy challenges with respect to CAV functions are summarised in Table 5. Designing DP mechanisms can mitigate the privacy risks for the challenges.

Table 5 Summary of identified privacy challenges for the CAV functions

CAV Function	Privacy Challenges
Route Planning	Revealing the desired route to the infrastructure to be able to receive the real-time optimal route plan
Participating in a Vehicle Platoon	Reveal sensitive information with either the infrastructure or other vehicles to be able to participate in a right platoon
HD/3D Map Updates	The temporal HD/3D Map updates should be trustworthy without revealing the publishers` privacy.
Vehicular Telemetry and Biometric Data Collection	Vehicular telemetric and biometric data is highly valuable for many parties. However, it should not reveal real-users privacy.

8 Conclusion

In this work, we have analysed the complete CAV systems regarding the privacy challenges arise from participating in ITS and CAV functions. We have also identified the approaches might be applied to mitigate such privacy challenges. However, it is also recognised that using such approaches can introduce a reduction in the efficacy of the CAV functions. This work can be supportive for academics and industry to investigate the privacy challenges in CAV systems, and to extend the research to narrow the knowledge gap in the area. It will open the door to design robust privacy-preserving mechanisms without jeopardising the efficacy of CAV functions.

9 Acknowledgements

This work is supported by the Alan Turing Institute under EPSRC grant EP/N510129/1 and partially funded through the UK Hub for Cyber Security of the Internet of Things, PETRAS, under grant (EP/N02334X/1).

10 References

[1] C. Maple, “Security and privacy in the internet of things,” *J. Cyber Policy*, vol. 2, no. 2, pp. 155–184, May 2017.

[2] H. Hasrouny, A. E. Samhat, C. Bassil, and A. Laouiti, “VANet security challenges and solutions: A survey,” *Veh. Commun.*, vol. 7, pp. 7–20, 2017.

[3] V. Vukadinovic *et al.*, “3GPP C-V2X and IEEE 802.11p for Vehicle-to-Vehicle communications in highway platooning scenarios,” *Ad Hoc Networks*, vol. 74, pp. 17–29, 2018.

[4] C. Schmittner, Z. Ma, C. Reyes, O. Dillinger, and P. Puschner, “Using SAE J3061 for Automotive Security Requirement Engineering,” vol. 1698, no. November 2018, 2016, pp. 157–170.

[5] T. Mekki, I. Jabri, A. Rachedi, and M. ben Jemaa, “Vehicular cloud networks: Challenges, architectures, and future directions,” *Veh. Commun.*, vol. 9, pp. 268–280, 2017.

[6] S. C. Woolley and P. N. Howard, “Computational Propaganda Worldwide: Executive Summary,” *Oxford Internet Institue*, no. 11, p. 36, 2017.

[7] M. Williams, J. R. C. Nurse, and S. Creese, “The perfect storm: The privacy paradox and the Internet-of-things,” *2016 11th Int. Conf. Availability, Reliab. Secur. ARES 2016*, pp. 644–652, 2016.

[8] A. Acquisti, L. Brandimarte, and G. Loewenstein, “Privacy and human behavior in the age of information,” *Science (80-.).*, vol. 347, no. 6221, pp. 509–515, 2015.

[9] M. Taddeo and L. Floridi, “The case for e-trust,” *Ethics Inf. Technol.*, vol. 13, no. 1, pp. 1–3, Mar. 2011.

[10] Z. Yan, P. Zhang, and A. V. Vasilakos, “A survey on trust management for Internet of Things,” *J. Netw. Comput. Appl.*, vol. 42, pp. 120–134, Jun. 2014.

[11] F. Qu, Z. Wu, F. Wang, and W. Cho, “A Security and Privacy Review of VANETs,” *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 6, pp. 2985–2996, 2015.

[12] P. Asuquo *et al.*, “Security and Privacy in Location-Based Services for Vehicular and Mobile Communications: An Overview, Challenges and Countermeasures,” *IEEE Internet Things J.*, vol. XX, no. XX, pp. 1–25, 2018.

[13] J. Ni, A. Zhang, X. Lin, and X. S. Shen, “Security, Privacy, and Fairness in Fog-Based Vehicular Crowdsensing,” *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 146–152, 2017.

[14] P. Samarati and L. Sweeney, “Protecting Privacy when Disclosing Information: k-Anonymity and its Enforcement Through Generalization and Suppression.,” *Proc IEEE Symp. Res. Secur. Priv.*, pp. 384–393, 1998.

[15] S. D. Warren and L. D. Brandeis, “The Right to Privacy,” *Harv. Law Rev.*, vol. 4, no. 5, p. 193, 1890.

[16] D. J. Solove, “A Taxonomy of Privacy,” *Univ. PA. Law Rev.*, vol. 154, no. 3, p. 477, Jan. 2006.

[17] L. Stefanick, *Controlling Knowledge: Freedom of Information and Privacy Protection in a Networked World*. 2011.

[18] C. Dwork, “Differential Privacy,” *Autom. Lang. Program. 33rd Int. Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II*, pp. 1–12, 2006.

[19] T. Dalenius, “Towards a methodology for statistical disclosure control,” *Stat. Tidskr.*, vol. 15, pp. 429–444, 1977.

[20] G. Danezis *et al.*, *Privacy and Data Protection by Design - from policy to engineering*, vol. abs/1501.0,

no. December. 2015.
[21] A. Bergström, "Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses," *Comput. Human Behav.*, vol. 53, pp. 419–426, 2015.
[22] A. F. Westin, "Privacy and Freedom," *Wash. Lee Law Rev.*, vol. 25, no. 1, p. 166, 1968.
[23] ETSI, "ETSI EN 302 665 Intelligent Transport Systems (ITS); Communications Architecture," *Etsi*, vol. 1, pp. 1–44, 2010.
[24] F. Cunha *et al.*, "Data communication in VANETs: Protocols, applications and challenges," *Ad Hoc Networks*, vol. 44, pp. 90–103, Jul. 2016.
[25] E. Hamida, H. Noura, and W. Znaidi, "Security of Cooperative Intelligent Transport Systems: Standards, Threats Analysis and Cryptographic Countermeasures," *Electronics*, vol. 4, no. 3, pp. 380–423, 2015.
[26] L. Zhu, F. R. Yu, Y. Wang, B. Ning, and T. Tang, "Big Data Analytics in Intelligent Transportation Systems: A Survey," *IEEE Trans. Intell. Transp. Syst.*, pp. 1–16, 2018.
[27] V. H. Le, J. den Hartog, and N. Zannone, "Security and privacy for innovative automotive applications: A survey," *Comput. Commun.*, vol. 132, no. September, pp. 17–41, 2018.
[28] D. Jia, K. Lu, J. Wang, X. Zhang, and X. Shen, "A survey on platoon-based vehicular cyber-physical systems," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 1, pp. 263–284, 2016.
[29] M. Amoozadeh, H. Deng, C. N. Chuah, H. M. Zhang, and D. Ghosal, "Platoon management with cooperative adaptive cruise control enabled by VANET," *Veh. Commun.*, vol. 2, no. 2, pp. 110–123, 2015.
[30] S. Karnouskos and F. Kerschbaum, "Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles," *Proc. IEEE*, 2017.
[31] J. A. Crossman, Hong Guo, Y. L. Murphey, and J. Cardillo, "Automotive signal fault diagnostics. I. Signal fault analysis, signal segmentation, feature extraction and quasi-optimal feature selection," *IEEE Trans. Veh. Technol.*, vol. 52, no. 4, pp. 1063–1075, Jul. 2003.
[32] R. Wang and S. M. Lukic, "Review of driving conditions prediction and driving style recognition based control algorithms for hybrid electric vehicles," in *2011 IEEE Vehicle Power and Propulsion Conference*, 2011, pp. 1–7.
[33] P. Taylor, N. Griffiths, A. Bhalerao, Z. Xu, A. Gelencser, and T. Popham, "Warwick-JLR driver monitoring dataset (DMD)," in *Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications - AutomotiveUI '15*, 2015, pp. 89–92.
[34] D. Brown and K. Bradshaw, "A multi-biometric feature-fusion framework for improved uni-modal and multi-modal human identification," *2016 IEEE Symp. Technol. Homel. Secur. HST 2016*, 2016.
[35] M. Bertoncello *et al.*, "Monetizing car data. New service business opportunities to create new customer benefits.," *Adv. Ind. McKinsey Co.*, no. September, 2016.
[36] L. Sweeney, "Achieving k-anonymity privacy protection using generalization and suppression," *Int. J. Uncertainty, Fuzziness Knowledge-Based Syst.*, vol. 10, no. 5, pp. 571–588, Oct. 2002.
[37] B. Fung, K. Wang, A. Fu, and P. Yu, *Introduction to Privacy-Preserving Data Publishing*, vol. 17. CRC Press, 2010.
[38] A. Machanavajjhala, J. Gehrke, and D. Kifer, "l-Diversity : Privacy Beyond k -Anonymity," *Proc. 22nd Int. Conf. Data Eng.*, vol. 1, pp. 1–36, 2006.
[39] N. Li, T. Li, and S. Venkatasubramanian, "t-Closeness: Privacy Beyond k-Anonymity and l-Diversity," in *2007 IEEE 23rd International Conference on Data Engineering*, 2007, no. 3, pp. 106–115.
[40] R. Chen, H. Li, A. K. Qin, S. P. Kasiviswanathan, and H. Jin, "Private spatial data aggregation in the local setting," *2016 IEEE 32nd Int. Conf. Data Eng. ICDE 2016*, pp. 289–300, 2016.
[41] A. Beimel, K. Nissim, and U. Stemmer, "Private Learning and Sanitization: Pure vs. Approximate Differential Privacy," *Theory Comput.*, vol. 12, no. 1, pp. 1–61, 2016.
[42] T. Zhang and Q. Zhu, "Distributed Privacy-Preserving Collaborative Intrusion Detection Systems for VANETs," *IEEE Trans. Signal Inf. Process. over Networks*, vol. 4, no. 1, pp. 148–161, 2018.
[43] F. Fioretto, C. Lee, and P. Van Hentenryck, "Constrained-Based Differential Privacy for Mobility Services," in *17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS 2018)*, 2018, pp. 1405–1413.
[44] Z. Zhou, Y. Qiao, L. Zhu, J. Guan, Y. Liu, and C. Xu, "Differential privacy-guaranteed trajectory community identification over vehicle ad-hoc networks," *Internet Technol. Lett.*, vol. 1, no. 3, p. e9, 2018.
[45] F. Kargl, A. Friedman, and R. Boreli, "Differential privacy in intelligent transportation systems," *6th ACM Conf. Secur. Priv. Wirel. Mob. Networks*, p. 107, 2013.
[46] B. Nelson and T. Olovsson, "Introducing Differential Privacy to the Automotive Domain: Opportunities and Challenges," in *2017 IEEE 86th Vehicular Technology Conference (VTC-Fall)*, 2017, pp. 1–7.
[47] J. Joy, V. Rabsatt, and M. Gerla, "Internet of Vehicles: Enabling safe, secure, and private vehicular crowdsourcing," *Internet Technol. Lett.*, vol. 1, no. 1, p. e16, 2018.
[48] L. Zhou, L. Yu, S. Du, H. Zhu, and C. Chen, "Achieving Differentially Private Location Privacy in Edge-assistant Connected Vehicles," *IEEE Internet Things J.*, vol. 4662, no. c, pp. 1–1, 2018.
[49] A. Wasef and X. Shen, "Efficient Group Signature Scheme Supporting Batch Verification for Securing Vehicular Networks," in *2010 IEEE International Conference on Communications*, 2010, pp. 1–5.

[50] Q. Zhang *et al.*, “OpenVDAP: An open vehicular data analytics platform for CAVs,” *Proc. - Int. Conf. Distrib. Comput. Syst.*, vol. 2018–July, pp. 1310–1320, 2018.
[51] J. Petit, F. Schaub, M. Feiri, and F. Kargl, “Pseudonym Schemes in Vehicular Networks: A Survey,” *IEEE Commun. Surv. Tutorials*, vol. 17, no. 1, pp. 228–255, 2015.
[52] A. Acar, Z. B. Celik, H. Aksu, A. S. Uluagac, and P. McDaniel, “Achieving Secure and Differentially Private Computations in Multiparty Settings,” pp. 49–59, 2017.
[53] A. Narayanan and V. Shmatikov, “How To Break Anonymity of the Netflix Prize Dataset,” 2006.
[54] J. Hsu *et al.*, “Differential privacy: An economic method for choosing epsilon,” *Proc. Comput. Secur. Found. Work.*, vol. 2014–Janua, pp. 398–410, 2014.
[55] C. Dwork and A. Smith, “Differential privacy for statistics: What we know and what we want to learn,” *J. Priv. Confidentiality*, vol. 1, no. 2, pp. 135–154, 2010.
[56] B. Avent, A. Korolova, D. Zeber, T. Hovden, and B. Livshits, “BLENDER: Enabling Local Search with a Hybrid Differential Privacy Model,” 2017.
[57] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, “Geo-Indistinguishability: Differential Privacy for Location-Based Systems,” pp. 901–914, 2012.
[58] T. Zhu, G. Li, W. Zhou, and P. S. Yu, “Differential Privacy and Applications,” vol. 69, 2017.