Buckle-up: Autonomous Vehicles Could Face Privacy Bumps in the Road Ahead

Ioannis Krontiris[1], Thanassis Giannetsos[2], Peter Schoo[1], and Frank Kargl[3]

[1] European Research Center, Huawei Technologies, Munich, Germany
{ioannis.krontiris, peter.schoo}@huawei.com
[2] Cyber Security, Department of Applied Mathematics and Computer Science, Technical University of Denmark
atgi@dtu.dk
[3] Institute of Distributed Systems, University of Ulm, Germany
frank.kargl@uni-ulm.de

Abstract. Autonomous vehicles, as part of the emerging Intelligent Transportation Systems (ITS), are positioned to transform the future of mobility — a change enabled by new on-board sensors, as well as the exchange of information between vehicles and between vehicles and transport infrastructure. This raises new and unique privacy considerations around what happens with the data. As the automotive industry becomes more data-driven, getting consumer privacy rights will become increasingly important for establishing trust and customer acceptance of this technology. In this paper we analyze what are the new privacy and data protection challenges that emerge in this domain and we put forth directions of research initiatives for overcoming these challenges. We build the discussion around legal compliance, identity management, in-vehicle data recording, and anonymization of vehicle data. We then debate on the advantages brought forth by emerging technologies (ranging from the intersection of distributed edge and fog computing to new 5G-enabled smart connectivity networks) and how such innovations can fulfill advanced privacy requirements in automotive industry.

Keywords: Privacy · Data Protection · Autonomous Driving · V2X Communication · In-vehicle Data Recording · Identity Management

1 Introduction

In the last years, there has been a lot of interest in the development of vehicles capable of driving autonomously. Autonomous vehicles (AVs) promise highly increased traffic safety and fuel efficiency, better use of the infrastructure, and the liberation of drivers to perform other tasks. For these reasons, autonomous driving may create a paradigm shift in the way people and goods are transported.

Connectivity and communication technology – V2V as well as V2X communication – is considered a key success factor paving the way for the successful implementation of autonomous driving functions. V2X communication enables two key features in AVs: cooperative sensing, which increases the sensing range by means of the mutual exchange of sensed data, and cooperative maneuvering, which enables a group of AVs to drive according to a common decision-making

strategy [25]. This connectivity between vehicles and between vehicles and transport infrastructure is expected to significantly improve road safety, traffic efficiency and comfort of driving, by helping the driver take the right decisions and adapt to the traffic situation [13].

AVs, however, go beyond just being connected vehicles. They are formally defined as those in which at least some aspects of safety-critical driving control occur without direct driver input. To achieve this goal, AVs require extensive data. Specifically, we are seeing the emergence of vehicles that feature an impressive array of sensors and on-board decision-making units capable of coping with an unprecedented amount of data. According to reports, sensors on AVs will generate data roughly between 1.4 TB/h and 19 TB/h [24].

As AV technology is still in its infancy, privacy aspects are not well addressed [35]. The capabilities of AVs pose new challenges especially on privacy protection, given the ubiquitous nature of capturing data in public and the ability to scale without additional infrastructure. Another aspect that complicates things even further is the fact that AVs capture data not only from users, but also from non-users (i.e. pedestrians walking outside the vehicle) with very limited possibilities to offer notice and choice about data practices.

Discussions on how to manage privacy risks from the legal perspective have been ongoing in the last years, but a lot of uncertainty still remains. The EU and governments in most countries have developed new regulations to control the access to, use and sharing of personal data, but these are not specific to AVs. Therefore, there is great need for clarifications specific to the context of Cooperative Intelligent Transport Systems (C-ITS) and AVs. This legal uncertainty about what can be done and what not burdens technological development in the automotive industry and so it is critical that it is addressed timely.

The AV ecosystem must also prioritize consumer expectations and trust. Since privacy is the baseline for trust into a system, it is a requirement for customer acceptance of a technology and, consequently, it is a key market enabler. A recent survey on the public opinion on automated driving reveals that there are worries on safety and privacy aspects of AVs [32]. Bloom et al. [6] also investigated people's conceptions of the sensing and analysis capabilities of AVs and found that scenarios such as tracking and identification caused overwhelming discomfort to people.

Contribution: In this paper we bring forward what we see as emerging privacy challenges in the autonomous driving domain, not only from the technological side, but also from the legal and policy landscape. We argue that the special characteristics of autonomous driving create an environment where current solutions fall short and we put forth directions of research initiatives for overcoming these limitations. We debate on the advantages brought forth by emerging technologies (ranging from the intersection of distributed edge and fog computing to new 5G-enabled smart connectivity networks) and how their adoption can be proven as an invaluable milestone for coping with the hurdles of current mechanisms that cannot capture the privacy and trust requirements of all involved stakeholders. We present our vision of how such innovations can fulfill these advanced

requirements and we highlight open issues and challenges that need to be taken into consideration at an early stage. We believe that advancing the discussion on these core pillars, in AVs, can be used to shape industry practices into developing and adopting privacy-respecting technologies, before deployment outpaces understanding of potential ramifications.

2 Data defined

To understand what kind of personal data are collected by AVs, we first need to look at the sensors that such vehicles are equipped with. Typical sensors include GPS for navigation, cameras located in the front, rear, left and right that give a 360° view of the car, and a multitude of ranging sensors like RADAR and light detection and ranging (LiDAR) for generating a 3D map of the environment. Data fusion integrates all this sensor data into an environmental model of a vehicle's surrounding that also includes detection of object types to distinguish, for example, between cars, pedestrians, bikers and solid obstacles.

Vehicles also collect data from their surrounding vehicles. As part of the European C-ITS technology, vehicles broadcast their speed, location and direction data using two types of messages, which are known as Cooperative Awareness Messages (CAM) and Decentralized Environmental Notification Messages (DENM) [10]. CAM messages are broadcast quasi-continuously (at 1-10 Hz) and they contain dynamic, kinematic data, as well as static information like dimensions of the vehicle. DENM messages are broadcast in addition to the CAM messages, but only upon the occurrence of specific events (like accidents) for urgent situations, and they contain location information about the event. Similar technology is standardized in other parts of the world.

Privacy concerns about vehicular communication have been raised already in the early 2000s [38]. More recently, the Data Protection and Privacy Working Group of the Cooperative Intelligent Transport Systems has issued an analysis, which makes it explicit that the broadcast CAM and DENM messages are personal data [9]. The reason for that is that even though they do not contain any unique identifiers, the data subject may be indirectly identifiable, either through the contained information like location data or the dimensions of the vehicle contained in the CAM messages, or through the PKI certificate, attached to both messages. The EU Commission [11] and the Art. 29 WP [4] also make it clear that data broadcast by vehicles qualify as personal data, as it relates to an identified or identifiable natural person.

Even though a number of stakeholders consider that most of the vehicle data should be treated as personal data by default [15], it is worth trying to list and categorize personal data based on how they are collected and disseminated in the system. This can make the difference between current vehicle technologies and AVs more clear and form the basis of analysing new privacy challenges. So first let us look at what kind of personal data are collected by AVs:

Data about ego vehicle and its passengers, for example:

- data on the drivers and passengers like name, address, account information, but also in-vehicle video and biometric data for the authentication of the driver (e.g., voice-, fingerprint-, video-and other types of authentication) or his monitoring (e.g., image processing for fatigue detection),
- data on personal devices of drivers and passengers like MAC addresses,
- trip information like start and end of trips,
- vehicle location data.

Data about vehicle-external entities, for example:

- license plates of surrounding vehicles,
- video recordings including identifying information like faces of pedestrians, bikers, etc.,
- sensor data from LIDAR, RADAR, etc. involving other persons,
- data received from other vehicles (location, etc.)

This multitude of data to be considered from a data protection perspective makes it challenging to effectively assess a C-ITS system in a data protection impact assessment. It can be helpful to classify data first and one way to categorize data is based on the way it can be accessed, for example whether it is broadcast to all, broadcast in a private network or can only be accessed from a stored source. The International Working Group on Data Protection in Telecommunications has suggested the following categories [28]:

- Data collected and processed by the vehicle, including information and entertainment systems built into the vehicle.
- Data exchanged between the vehicle and personal devices connected to it,
- Data exchanged between the vehicle and external entities (e.g., infrastructure managers, vehicle manufacturers, insurance companies, car repairers).
- Data broadcast to surrounding vehicles and infrastructure entities to enable C-ITS.

Another way to categorize data is to look at the source of the data, as proposed in Table 1.

3 Bumpy Road Ahead

While privacy protection for C-ITS has been investigated for more than a decade and solutions like changing pseudonyms have been included into standards, we think that autonomous driving will create new challenges or aggravate existing ones. Therefore, in this paper we want to discuss those emerging privacy challenges and argue on the shortcomings of current practices from industry and research community to address these challenges in adequate ways.

Table 1. Data related to autonomous vehicles.

Kind of data	Example how data is handled
Sensor data (sensors, radars, Lidar)	High-bandwidth sensor data for object avoidance and mapping, and infrared thermal imaging. Data acquired and processed on-board. Under some conditions part of the data may be send off-board for training the machine learning system.
Video Recording (exterior)	Capture high-bandwidth images of vehicles and parties external to the vehicle. Identify external parties and number-plates of other vehicles. Data acquired and processed on-board. Under some conditions part of the data may be sent off-board for training the machine learning system.
Video Recording (interior)	Monitor driver alertness and occupant behaviour. Data acquired on-board and can stay there.
Biometric, biological or health data	Monitor driver alertness and behavior. Recognise drivers and occupants through fingerprints or facial recognition
Crash-related data	Input data from the vehicle in the seconds before and during the crash stored in Event data recorder. Data collected and processed by the vehicle.
V2X Communication data	Enable awareness driving through CAM and DENM in EU or BSM in US, sensing driving through Collective Perception Message (CPM) and cooperative automated driving trough Maneuver Coordination Message (MCM) and Platooning Control Message (PCM).

3.1 World-wide Legal Compliance

OEMs produce for a world-wide market. Many countries are updating and enhancing their data protection laws, as evidenced by the European GDPR, the Japanese Act on the Protection of Personal Information (APPI), or the California Consumer Privacy Act (short CCPA) that all were created or revised in the last years. This creates the problem of how to design and develop C-ITS systems that are compliant with those fast changing and in parts diverse regimes. Al-Momani et al. [2] provide a discussion of different legal data protection regimes world-wide and the consequences when developing automotive services for markets worldwide.

3.2 Legal Bases for Processing Personal Data

As a general principle, each company processing personal data as a controller needs a lawful basis to do so. One such lawful basis is informed consent, where the individuals affected about the intended uses of their personal data get informed about this and their consent to such processing is obtained. However, for selling and offering services around AVs, consent is not the only option. Article 6(1) specifies several other options that can make processing of personal data lawful, for example, if required for fulfillment of a contractual obligation with that

person. Regarding AVs, the legal discussions are still ongoing as to which of these options apply in which case, since the situation is complex with many different players involved, each having different purposes for the data collected.

The case of legal basis for processing CAM and DENM messages in the context of C-ITS and connected car is perhaps an indicative example. The Data Protection WG of the C-ITS Platform [9] has analysed thoroughly each of the above options for legal basis and it has given comments on the feasibility of each one. Regarding informed consent, it makes it clear that this form of legal basis is simply impossible in practice. The Art. 29 WP [4] has analysed this further and has given several reasons for the difficulty of implementing consent as legal basis. They mainly have to do with the fact that car owners and car users have to be treated separately and that broadcasting nature of the communication makes it impossible to establish a mutual recognition mechanism between the data subject (sender) and controller (recipient). On the other side, the recent Guidelines published by the European Data Protection Board (EDPB) communicate the view that consent should generally be the legal basis for the processing of personal data in relation to connected vehicles [14].

In the context of cameras, video data in which individuals are recognizable amounts to personal data for GDPR purposes and so a valid legal basis is required for data processing activities. Under Article 89 (and Recital 159) of the GDPR, there is an exemption that permits the processing of personal data for scientific research purposes which is even strengthened by some national or state-level regulation. The Bavarian Data Protection Authority has indicated that it considers that the use of dashcams for the purposes of research and development of autonomous vehicles could fall within this exemption [5] and §13 in the state-level data protection regulation in Baden-Württemberg points in a similar direction [33]. However, it should be noted that there is no broader consensus on a European level at this point.

So there is a need for clear guidance to help controllers and technology designers determine on what lawful basis for processing to rely on and implement the corresponding solutions. This should happen based on data categories as outlined above. Relying on consent alone may prove challenging in the context of connected and autonomous driving. Of particular interest to technology designers is the fact that provisioning of consent requests must be in "intelligible and easily accessible form, using clear and plain language" according to GDPR. This touches the issue of usability of privacy notices and controls, which even though it has been studied extensively, few ideas exist in the automotive domain. In particular for safety messaging, establishing a legal basis for processing may be the more promising approach, but requires the legislation to become active.

3.3 Transparent and Interpretable Processing

As emphasized by Art. 29 WP [4], users need to be fully aware of the scope of the processing of the data they broadcast through their vehicles. Who receives these data (e.g. other vehicles, OEMs, road managers, etc.) and how they process these data should also be transparent to the data subjects. The Data Protection

and Privacy Commissioners also urge the involved parties to give data subjects comprehensive information as to what data is collected and processed in the deployment of connected vehicles, for what purposes and by whom [27].

There are currently only limited possibilities to interact with the data subject within the vehicle and provide sufficient and appropriate information about the handling of personal data. These limitations are mostly due to the small displays (such as the head unit in the vehicle) or mobile apps, which make it hard to transparently provide the user with all relevant information. One particular OEM currently shows a 15 page long privacy policy on a 20 cm screen. So it is a challenge for future research to create new user interface design as well as privacy policy formatting and make the collection, use, and sharing of information more comprehensive to the data subject.

As future mobility with autonomous mobility is expected to also show a much higher degree of shared mobility, this also means that passengers will likely change cars very frequently. Reconfiguring privacy-settings and consenting in each vehicle anew with completely different UIs is not practical, so we need a higher degree of standardization of privacy settings among OEMs and maybe even a privacy profile portability where users can carry their settings with them, e.g., on a smartphone.

Outside the vehicle, a company deploying cameras typically has no direct relationship with the individuals who may pass through the dashcam's field of view, which makes it more challenging to provide those individuals with the required information. The Guidelines suggest the use of a "layered" approach, with the most important information displayed on a highly visible sign (e.g., a sticker on the outside of the vehicle) alerting individuals to the fact that a dashcam is being used, and providing a means of obtaining further information (e.g., using a QR code that individuals can scan with a smartphone, and that links to an online privacy notice setting out the required information).

3.4 Explainability of AI

Related to interpretability and transparency is the notion of explainable AI. As the level of automation is constantly increasing via the use of state-of-the-art AI solutions, such systems become more difficult to understand for the user and there is no transparency around algorithmic decision making. In order to address this problem, the EU's GDPR has added the "right to an explanation" to the policy, highlighting the importance of human-understandable interpretations derived from machine decisions. Further, the regulation is requiring firms to provide data subjects with "meaningful information about the logic involved" in "concise, intelligible and easily accessible" forms [21].

However, legally the GDPR does not seem to fully guarantee or grant "right to an explanation", since it is not part of the regulation itself but rather appears only in recital 71 which has no binding nature [41]. Although it is not yet clear how these legal requirements will be implemented in practice, it is unlikely that there will be a one-size fits all technical approach to explainability, given the wide range of AI applications. Any solution should be attuned to what each AI

technology is expected to accomplish in the context where it is applied and what are the possibilities to reveal meaningful information. In any case, one can be sure that transparency aspects will gain in importance as AI decisions will affect our daily lives more and more.

3.5 Identity Management

Part of the discussion on privacy for connected and AVs concern the security of the messages exchanged between vehicles, i.e. the CAM and DENM messages. It was suggested very early to use digital certificates to authenticate messages in vehicular communications and prevent an attacker from injecting false messages [20]. Certificate do not contain any information that links them to a particular vehicle or owner, in order to protect the privacy of individuals from location tracking. Instead, vehicles are assigned pseudonyms, which is embedded in certificates, in order to preserve anonymity [18].

Aiming to cope with the management of pseudonym certificates, many proposals have appeared in the literature for creating a Vehicular Public Key Infrastructure (VPKI) (for a survey, see [30]). Two prominent examples are the Security Credential Management System (SCMS) [7], which is a product of vehicle OEM consortia and the US Department of Transport (USDOT), and the European Cooperative-ITS Certificate Management System (CCMS), developed by CEN and ETSI with support from the Commission [12].

Both systems envisage a technical and organizational separation of duties between different PKI authorities to cope with internal attackers, ensuring that "no single entity" in the architecture can track a vehicle across space and time, unless it colludes with one or more entities. This is in line with other research on VPKI structures that are inherently resilient to such collusion inside the PKI [16, 17]. However, since it would be a costly solution to realize this, in practice it is not precluded that multiple authorities, operating these entities, are under the same organizational umbrella. For example, USDOT describes removal of certain separations of SCMS functions, which may now reside in the same organization, while the responsibility is passed to the SCMS Manager to decide on the rules for governance/policy of separation [40]. Note that the SCMS manager is expected to be an industry-wide coalition of stakeholders. Another reason that justifies the shift of onus to the SCMS Manager is the inclusion of the Mobile Network Operator (MNO) in the ecosystem, which leads to the re-evaluation of the risk assessment and allows a variety of simplifications, as pointed out by 5GAA [1].

However, it is clear that removing the "no single entity" constraint does weaken the privacy protection to some extent, and hence may increase the risk of vehicle tracking by internal entities. This therefore has to be considered carefully when it comes to actual operation of a V2X PKI, taking all risks and mitigation into account. Concerning PKI operation, organizational separation based on technical means still offers higher privacy protection assurances. In general, different parties or authorities inside the V2X PKI ecosystem can possibly collude together to compromise privacy and track a vehicle, despite other policies. A frequent assumption in this case is that all entities in the system are

honest and don't collude with each other. Thus, the question remains, how to establish and maintain this federated trust and experience.

3.6 Anonymization of Vehicle Data

There are several cases where anonymization of vehicle data is advised or even legally required. In the context of AVs, huge amounts of video data are being collected with the goal of creating environmental models to train the intelligent algorithms or to validate complex autonomous driving functionality. Images or videos recorded in public areas may contain personal data such as license plates and people's faces. AVs cannot give all pedestrians and drivers they encounter notice and choice. So another permitted solution to respect privacy rights is to anonymize the recorded data immediately, so that no conclusion about personal information can be drawn.

Early methods have been relying on obfuscation with a solid colored box, pixelization, random pixel shuffling, Gaussian blur and distortion. For example, Grosselfinger et al. have presented an architecture for automatic multimodal video data anonymization to ensure data protection [22]. Schnabel et al. [39] evaluated several techniques and concluded that anonymization of personal data in the training set can impact the detection of vehicles at various degrees. However one would expect that anonymization would have a different impact on detection performance, depending how important the region we target is for feature learning. For example, perhaps blurring license plates could have a different impact on the performance of a car detector, than the impact of blurring faces on the human detector. More extensive experimental evaluation is definitely needed in this area. Another approach is to replace faces or number plates in the video with generated ones, in order to de-identify subjects in images or videos, while preserving non-identity-related aspects of the data and consequently enabling better data utilisation.

3.7 In-vehicle Data Recording, Storage and Access Management

Aiming to support crash reconstruction or accident investigations, automotive products have been developed collecting and recording event data. With advent of autonomous driving, forensic capabilities require massive enhancements: in addition to traditional low-bandwidth sensor data, now also data that contribute to driving and maneuver decisions are required for crash reconstruction analysis. High-bandwidth data is required to be stored, if the operation of machine learning based autonomous decision making shall be reconstructible and the execution of the computing decision that controlled the vehicle shall be forensically analyzed. The reconstruction of Uber's Elaine Herzberg case presents a first very relevant example [23].

For sure, event data recording can include information that is related to an identified or identifiable natural person. It is a matter of the (i) purpose for which the event data recording is performed that determines the privacy

requirements, and *per se* event data recording is a (ii) subset of all information managed on-board. The discussion in this section focuses on event data recording for reconstruction or clarification of accidents, i.e. low-bandwidth and high-bandwidth data collection on-board that serves the forensic analysis. The technical challenge is to store the necessary data in time and according to legal requirements. Based on existing US and upcoming EU regulatory requirements, any data recording beyond on-board collection that fulfills other purposes than forensics will have different security and privacy goals. For example, if recorded data is condensed to optimize storage requirements and regularly send into the OEMs cloud infrastructure [37].

Whereas historically some of the sources that contribute to the event recording are distributed on-board of a vehicle, we see the future tendency to have one physical device and one interface only.

Table 2. Forensic Event Data Recorder Requirements.

Forensic Goals	**Data Protection Goals**
- Recording necessities (specific to vehicle capabilities)	- Data availability
- Authenticity of data	- Data integrity
- Data consistency including timing correctness	- Access control at interface when reading recorded data

The event data stored on such a device is neither available via a public interface nor is the physical interface easily accessible. Sometimes it is even necessary to disassemble the interior of a vehicle to access the device. All this indicates that an investigation for crash reconstruction or accident analysis is left to specialists only. It is performed in case evidence is collected that have to fulfill modern court standards, when it comes to legal proceedings (*cf.* Table 2). Against this background, of course GDPR will play a role. However, it may be that higher-priority law, including court orders, supersedes data protection rights.

4 Bridging Security Management Silos to Boost Operational Assurance and Functional Safety in AVs

Seeking to design successful secure and privacy-preserving architectures for AVs, one has to cater for the open challenges we discussed in the previous sections. The security, interoperability and connectivity in a dynamic network of vehicles, gateways, services and applications across operations technology and information technology stakeholders requires strategic rethinking of policies and processes in the context of cyber-security, privacy and trust establishment without impeding the strict safety requirements of such environments.

As a consequence, the current trend is to understand AVs inherently and increasingly as Federated Safety Critical Systems designed, implemented, operated and owned by multiple tenants capable of providing mixed-criticality services

with different security, privacy goals, requirements and priorities. Towards this direction, novel development paradigms need to become prominent leveraging the benefits of emerging technologies such as cloud-, edge-, and fog-computing, and their intersection, that are based on the microservice concept [29].

A high-level view of such an approach is depicted in Figure 1. Edge and fog computing nodes coexist in a 5G fronthaul-backhaul infrastructure and support the mixed-criticality microservices running either in the back-end cloud infrastructure or closer to the edge. The edge/fog level is composed by a diverse type of nodes and infrastructures with different processing capabilities, split in general into autonomous vehicles and fog-based service delivery nodes (e.g., application servers, content delivery and data storage nodes).

To achieve high scalability and effective agility levels, applications need to be decomposed into a mesh of "cloud-native" and "edge-running" microservices, each one with specific security, privacy, trust and safety objectives packaged on independent virtual execution environments. For instance, consider a safety-critical microservice, such as collision avoidance or accident reconstruction, with much more strict performance and trust boundaries than a microservice dedicated to traffic prediction or infotainment. As autonomous driving generates high-bandwidth data, it is a valid question if such data shall stay on board or is shared with instances outside the AV. A central entity, named "application orchestrator", is responsible for realizing this application business logic by properly managing the lifecycle and interconnection of such microservices over cloud and edge (i.e., AV) resources. Essentially, managing the accelerated offloading of the time and resource intensive (non-safety critical) microservices to the back-end cloud infrastructure, thus, reducing the load of the edge AVs (to be allocated directly towards the execution of only those safety-critical microservices that need to operate at the highest trust level) which in turn will allow the entire system to meet the strict requirements, as have been identified by the standards.

Depending on the complexity of the safety and security, privacy assurances that a microservice needs to adhere to, such an architecture allows for a smart, flexible slicing of the underlying infrastructure that can dynamically adapt to changing requirements (be it bandwidth, network, security, privacy and trust resources) of safety-critical applications whilst enabling network segmentation, thus, supporting real-time and latency free security and safety utilities. This milestone, however, can only be achieved through the adoption of 5G-enabled ecosystems capable of reducing the end-to-end latency between the cloud resources and the connected edge AVs to enable real-time security and safety data offloading procedures outside the AV.

Particularly with respect to safety and security, microservices must be enabled to make and prove statements about their state and actions so that other microservices can align their actions appropriately and an overall system state can be assessed where security policies can be evaluated and enforced.

For instance, single ECUs in a vehicle may be able to remotely attest to other devices that they are in an untampered state of integrity and have up-to-date and valid sensor input available on which it bases its decisions. Based on such

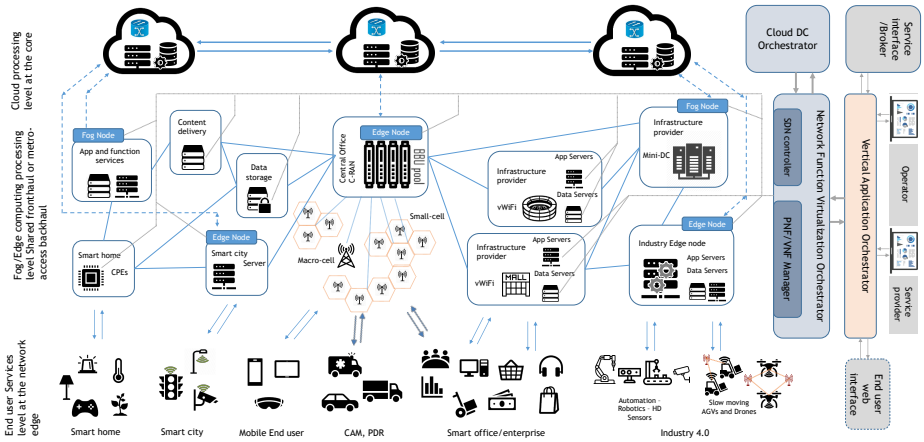

Fig. 1. Envisioned 5G Ecosystem for Autonomous Vehicles with intelligent network control and microservice deployment

a proof, another ECU may then decide (based on a security policy) that it can accept commands from the earlier ECU without risking safety and security of the overall system. This can be continued to ensure safety and security of an overall car and then of systems of cars communicating via Car-2-Car communication.

This goes substantially beyond simple authorization schemes telling who may access whom, but will require understanding of semantics of requests and chains of effects throughout the system and an analysis both statically at design-time and dynamically during run-time. It requires strong attestation services, targeting both the software and hardware layers and covering all phases of an AV's execution; from the trusted boot and integrity measurement, enabling the generation of static, boot-time or load-time evidence of the AV's components correct configuration (Configuration Integrity Verification (CIV)), to the run-time behavioral attestation of those safety-critical components of an AV providing strong guarantees on the correctness of the control- and information-flow properties [31], thus, enhancing the performance and scalability when composing secure AVs from potentially insecure components.

On a similar note, we also argue that the pressing need for establishing federated trust between services and devices cannot be solely secured, in such architectures, with common centralized solutions like PKIs. What is needed are decentralized solutions capable of (partially) shifting trust from the back-end infrastructure to the edge (i.e., vehicles) so as to reduce the vector of entities for which we want to make sound statements in terms of their configuration, security settings and trustworthiness.

For example, Camenisch et al. [8] introduced recently a solution based on the novel concept of zone encryption incorporating dynamic group signatures with attributes. This allows vehicles to generate unlimited pseudonyms locally with negligible credential download-and-storage costs. Similarly, Whitefield et al. [42]

suggest the use of Direct Anonymous Attestation (DAA) algorithms and trusted computing technologies as an enabler for more decentralized approaches, where trust establishing is shifted from the back-end infrastructure to the edge [19].

In the same line of research of investigating the integration of trusted computing technologies, another development would be to leverage software-based Trusted Execution Environments (TEEs), such as the Intel SGX [26] and the ARM TrustZone [3]. This could allow to confine processing of personal data within a secure enclave that is verified by remote attestation to be in a certified process that will not process personal data outside of the declared purpose.

Another direction towards a decentralized environment would be to investigate advanced AI techniques like federated learning [36], which enables developers to train based on the shared models on their decentralized devices or servers with the local dataset. Compared to traditional centralized machine learning techniques, federated learning reduces privacy concerns by maintaining data in local servers and sharing model updates, e.g., gradient information, instead of the raw data. However several challenges remain to make this technique more privacy preserving [34].

5 Conclusions

To conclude, we believe that a critical part of any effort to achieve consumer acceptance of AVs will be assuring consumers that the involved technologies do not pose a significant threat to privacy and have been designed to help protect against vehicle tracking by any government or company participating in the ecosystem. Therefore we see the need to address several remaining data protection issues, such as the lawful basis for processing, on a regulation and policy making level, in order to offer more clear guidance to the industry and enable them to progress faster. On the same level, there is also great need for more harmonization efforts between different legal data protection regimes world-wide. On a technical level, we identified the need to move towards scalable and decentralized solutions, eliminating the need for federated infrastructure trust. We discussed how this can be done by adopting emerging technologies, such as the intersection of distributed edge and fog computing with the new 5G-enabled smart connectivity networks, decentralized PKI architectures, trusted computing technologies and privacy-preserving machine learning in automotive contexts. While several research challenges remain open, we hope that the adoption of these technologies can be an invaluable milestone for coping with the hurdles of current mechanisms and shape the foundation for more privacy-preserving practices in the autonomous vehicle industry.

References

1. 5G Automotive Association: 5GAA Efficient Security Provisioning System. White Paper (2020)
2. Al-Momani, A., Kargl, F., Bösch, C.: A Comparison of Data Protection Regulations for Automotive Systems. Presented at CPDP 2020 (January 2020)

3. ARM Ltd.: GlobalPlatform based Trusted Execution Environment and TrustZone® Ready. White Paper (October 2013)
4. Article 29: Opinion 03/2017 on Processing personal data in the context of Cooperative Intelligent Transport Systems (C-ITS). Document (October 2017)
5. Bayerisches Landesamt für Datenschutzaufsicht: 8. Tätigkeitsbericht des Bayerischen Landesamts für Datenschutzaufsicht für die Jahre 2017 und 2018. Report
6. Bloom, C., Tan, J., Ramjohn, J., Bauer, L.: Self-driving cars and data collection: Privacy perceptions of networked autonomous vehicles. In: Thirteenth Symposium on Usable Privacy and Security (SOUPS 2017). Santa Clara, CA (2017)
7. Brecht, B., Therriault, D., Weimerskirch, A., Whyte, W., Kumar, V., Hehn, T., Goudy, R.: A Security Credential Management System for V2X Communications. IEEE Transactions on Intelligent Transportation Systems **19**(12) (Dec 2018)
8. Camenisch, J., Drijvers, M., Lehmann, A., Neven, G., Towa, P.: Zone encryption with anonymous authentication for V2V communication. In: 5th IEEE European Symposium on Security and Privacy (September 2020)
9. Data Protection and Privacy Working Group of the C-ITS Platform: Processing personal data in the context of C-ITS. Document (March 2017)
10. ETSI: Intelligent Transport Systems (ITS); Security; Security Header and Certificate. Technical specification TS 103 097 (2017)
11. EU Commission: A European strategy on Cooperative Intelligent Transport Systems, a milestone towards cooperative, connected and automated mobility. COM(2016) 766 final (November 2016)
12. European Commission: Certificate Policy for Deployment and Operation of European Cooperative Intelligent Transport Systems (C-ITS). Result of C-ITS Platform Phase II (2018)
13. European Commission: Intelligent transport systems - Cooperative, connected and automated mobility (CCAM). `https://ec.europa.eu/transport/themes/its/c-its_en` (accessed January 29, 2020)
14. European Data Protection Board: Guidelines 1/2020 on processing personal data in the context of connected vehicles and mobility related applications. Guidelines
15. Federation Internationale de l'Automobile: What EU legislation says about car data - Legal Memorandum on connected vehicles and data. Memorandum (2017)
16. Förster, D., Kargl, F., Löhr, H.: PUCA: A pseudonym scheme with user-controlled anonymity for vehicular ad-hoc networks (VANET). In: Vehicular Networking Conference (VNC), 2014 IEEE. pp. 25–32. Paderborn, Germany (Dec 2014)
17. Förster, D., Löhr, H., Zibuschka, J., Kargl, F.: REWIRE – Revocation Without Resolution: A Privacy-Friendly Revocation Mechanism for Vehicular Ad-Hoc Networks. In: Trust and Trustworthy Computing (2015)
18. Gerlach, M.: Assessing and Improving Privacy in VANETs. In: Proceedings of the 4th Workshop on Embedded Security in Cars (ESCAR) (2006)
19. Giannetsos, T., Krontiris, I.: Securing V2X Communications for the Future: Can PKI Systems Offer the Answer? In: Proceedings of the 14th International Conference on Availability, Reliability and Security. ARES '19 (2019)
20. Gollan, L., Meinel, C.: Digital Signatures For Automobiles?! In: Proceedings of Systemics, Cybernetics and Informatics (SCI). pp. 1–5 (July 2002)
21. Goodman, B., Flaxman, S.: European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation". AI Magazine **38**(3), 50–57 (Oct 2017)
22. Grosselfinger, A.K., Münch, D., Arens, M.: An architecture for automatic multimodal video data anonymization to ensure data protection. In: Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies III. vol. 11166, pp. 206 – 217 (October 2019)
23. Harris, M.: NTSB Investigation Into Deadly Uber Self-Driving Car Crash Reveals Lax Attitude Toward Safety - IEEE Spectrum. `https://spectrum.ieee.org/cars-that-think/transportation/self-driving/ntsb-investigation-into-deadly-uber-selfdriving-car-crash-reveals-lax-attitude-toward-safety` (accessed September 2020)

24. Heinrich, S.: Flash memory in the emerging age of autonomy. In: Flash Memory Summit 2017 Proceedings (August 2017)
25. Hobert, L., Festag, A., Llatser, I., Altomare, L., Visintainer, F., Kovacs, A.: Enhancements of v2x communication in support of cooperative autonomous driving. IEEE Communications Magazine **53**(12), 64–70 (Dec 2015)
26. Intel Corp: Intel Software Guard Extensions. `http://www.intel.com/content/www/us/en/processors/architectures-software-developer-manuals.html` (accessed June 2020)
27. International Working Group on Data Protection in Telecommunications: Resolution on Data Protection in Automated and Connected Vehicles. Resolution from the 39th International Conference of Data Protection and Privacy Commissioners, Hong Kong (September 2017)
28. International Working Group on Data Protection in Telecommunications: Connected Vehicles. Working Paper from the 63rd meeting, 9-10 April 2018, Budapest, Hungary (April 2018)
29. Jamshidi, P., Pahl, C., Mendonça, N.C., Lewis, J., Tilkov, S.: Microservices: The journey so far and challenges ahead. IEEE Software **35**(3), 24–35 (2018)
30. Khodaei, M., Papadimitratos, P.: The key to intelligent transportation: Identity and credential management in vehicular communication systems. IEEE Vehicular Technology Magazine **10**(4), 63–69 (Dec 2015)
31. Koutroumpouchos, N., Ntantogian, C., Menesidou, S.A., Liang, K., Gouvas, P., Xenakis, C., Giannetsos, T.: Secure edge computing with lightweight control-flow property-based attestation. 2019 IEEE Conference on Network Softwarization (NetSoft) pp. 84–92 (2019)
32. Kyriakidis, M., Happee, R., de Winter, J.: Public opinion on automated driving: Results of an international questionnaire among 5000 respondents. Transportation Research Part F: Traffic Psychology and Behaviour **32**, 127–140 (2015)
33. Landesbeauftragter für den Datenschutz und die Informationsfreiheit Baden-Württemberg: 34. Tätigkeitsbericht 2018. Report (2018)
34. Li, T., Sahu, A.K., Talwalkar, A., Smith, V.: Federated learning: Challenges, methods, and future directions. IEEE Signal Processing Magazine **37**(3), 50–60 (2020)
35. Lu, N., Cheng, N., Zhang, N., Shen, X., Mark, J.W.: Connected vehicles: Solutions and challenges. IEEE Internet of Things Journal **1**(4), 289–299 (Aug 2014)
36. McMahan, H.B., Moore, E., Ramage, D., y Arcas, B.A.: Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS) (2017)
37. Patsakis, C., Solanas, A.: Privacy-Aware Event Data Recorders: Cryptography Meets the Automotive Industry Again. IEEE Communications Magazine **51**(12), 122–128 (December 2013)
38. Raya, M., Hubaux, J.P.: The Security of Vehicular Ad Hoc Networks. In: Proceedings of the 3rd ACM Workshop on Security of Ad Hoc and Sensor Networks. p. 11–21. SASN '05 (2005)
39. Schnabel, L., Matzka, S., Stellmacher, M., Pätzold, M., Matthes, E.: Impact of anonymization on vehicle detector performance. In: Second International Conference on Artificial Intelligence for Industries (AI4I). pp. 30–34 (2019)
40. U.S. Department of Transportation, National Highway Traffic Safety Administration: Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application. DOT HS 812 014 (2014)
41. Wachter, S., Mittelstadt, B., Floridi, L.: Why a right to explanation of automated decision-making does not exist in the general data protection regulation. International Data Privacy Law **7**(2), 76–99 (2017)
42. Whitefield, J., Chen, L., Giannetsos, T., Schneider, S., Treharne, H.: Privacy-enhanced capabilities for VANETs using direct anonymous attestation. In: 2017 IEEE Vehicular Networking Conference (VNC). pp. 123–130 (Nov 2017)