

Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles

S. Karnouskos and F. Kerschbaum

Proceedings of the IEEE
(2017)

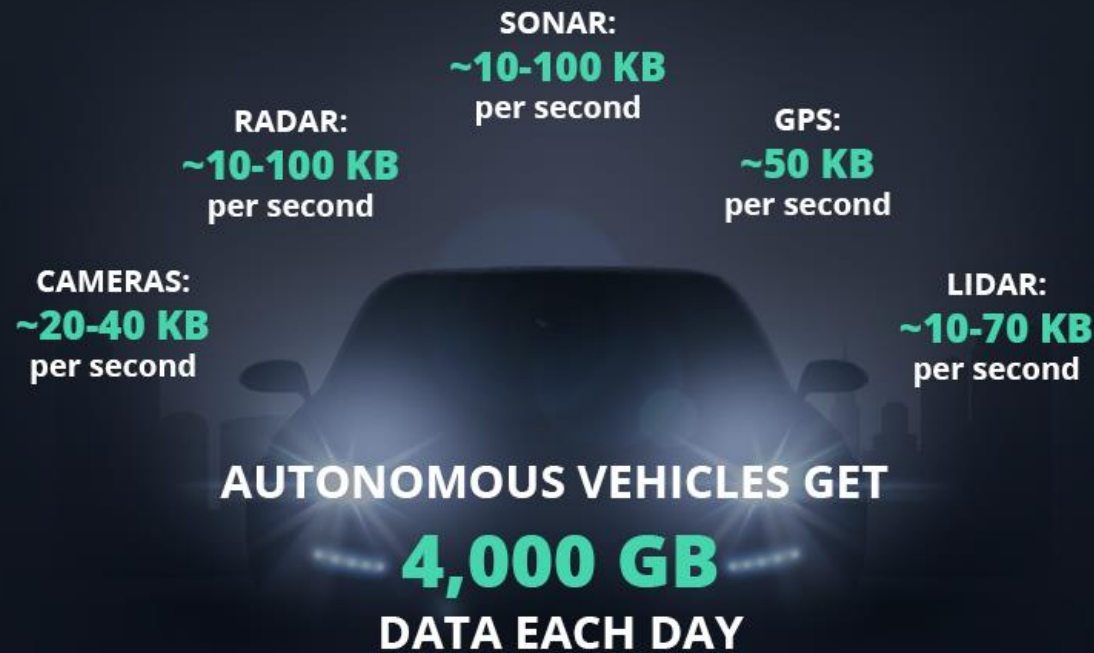
What does the paper address?

This paper examines the privacy and data integrity issues in the operation of fleets of cooperating, autonomous vehicles.

This work puts forward the hypothesis that it is feasible to ensure integrity, and hence safety, while preserving privacy in the emerging hyperconnected vehicle scenarios.

An exemplary case study on real-time vehicle interactions pertaining to map updates exemplifies the combination of privacy-enhancing technologies with integrity-protecting mechanisms.

What is the hyperconnected vehicle?



✕ An autonomous electric vehicle (EV), such as a car, that is capable of V2X communications and actively interacts with its surroundings and participates implicitly or explicitly in its complex cross-domain processes, e.g., within a smart city.

✕ A hyperconnected vehicle poses an ecosystem that includes the additional sensors and devices brought by its users (e.g., infotainment system, mobile phones, GPS driving systems, cameras) as well as the devices that explicitly or implicitly interact with the vehicle, e.g., roadside units (RSUs), and other vehicles.

What is the hyperconnected vehicle?



Data generated by the vehicle directly or indirectly as it interacts with other cyber-physical entities and services are gaining importance.






The increased number of sensors in hyperconnected cars generate a significant amount of data, for instance 25 GB/h, that can be assessed and used to take informed decisions.






In this context, the vehicle of the future can be seen as both a data platform, hosting the data generated by the vehicle itself, as well as a service platform, mediating access to that data (potentially in collaboration with cloud-based services).

A key question that emerges is how to reconcile the security goals of multistakeholder interactions that result due to the hyperconnectivity of the vehicle, i.e., preserve privacy of the vehicle's passengers and integrity of the infrastructure as whole.

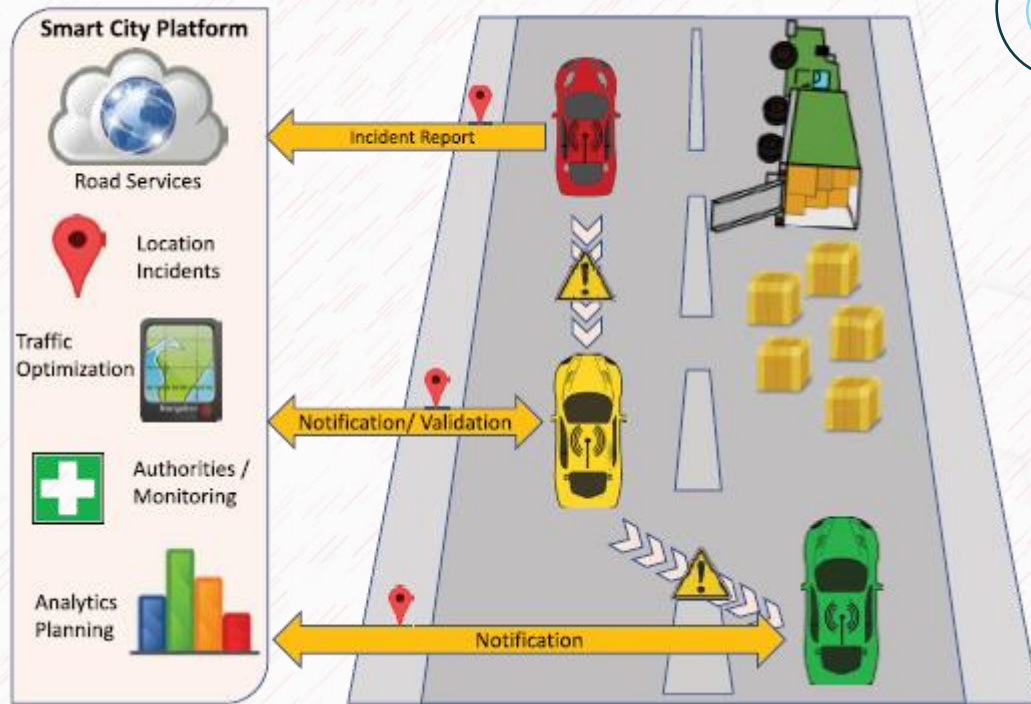
What's the conflict between integrity and privacy?

-  Integrity prevents unauthorized users from modifying or forging data and guarantees that all data are reliable, accurate, consistent, and of verifiable quality.
-  Privacy guarantees that the information acquired is appropriately utilized as intended, and while third parties can process it they ought not to derive intelligence from it.
-  On one hand, in multistakeholder interactions the integrity of the sensed data is key toward ensuring a safe and stable system.

What's the conflict between integrity and privacy?

-  Verifying the integrity of sensed data requires checking the possibly remote sensor's reading with contextual readings of other sensors.
-  However, the straightforward approach requires disclosing both readings and running a potentially complex statistical analysis.
-  On the other hand, privacy is founded on the principle of data minimization. While it is easy to disclose additional information in privacy-compliant system, the reverse is very difficult.

Example Scenario: Real-Time Map Updates



High-precision maps can be used to assist the navigation of the vehicle as well as other auxiliary vehicle-related services (e.g., route optimization based on the state of charge of the battery of the EV and the available charging stations in the area).



Integrity plays a pivotal role! The information that has to be present in the map must be highly credible and verifiable.



The dilemma posed is how to update the maps and their respective info with data coming from the field (i.e., the hyperconnected vehicles), which, in general, cannot be assumed as trusted entities.

What's the use for this example scenario?

In-vehicle detection:

It is not guaranteed that all vehicles will have the same level of sophistication, capabilities, or quick reactions, etc.



Vehicle-to-vehicle:

It depends on the density of the vehicles, and is susceptible to, e.g., misinformation from third parties (other vehicles).

Vehicle-to-infrastructure:

It enables the information dissemination from a trusted point of distribution, but as can be assumed they are susceptible to attacks at various levels, where security, trust, and privacy play a pivotal role.

Security Objectives and Challenges of Hyperconnected Vehicles

-  CPSs are multistakeholder, distributed systems and hence diverse stakeholders have different (security) objectives. Often these security objectives are in conflict and need to be balanced against each other.
-  The conflicting objectives in hyperconnected vehicles include **privacy**, **service offerings**, **data quality** and **integrity**, **spontaneous interactions**, and **safety**.

Privacy



Privacy in hyperconnected vehicle refers to the passenger's ability to control the use and storage of the data collected by the vehicle.



The vehicle is moving toward a general platform that is mobile, monitors the environment (video, sound, sensors), and collects detailed data that otherwise might not be shared, which effectively renders it to a potential privacy infringer.



The owner of the vehicle therefore has a vested interest and in many jurisdictions the right to control these data.



The authors envision technical measures that ensure privacy that can be verified by the user. An important, if not the most important, privacy principle in this context is that of data minimization which allows the user to retain control of all collected data.

Service Offering



Service offering refers to the ecosystem of electronic services around a hyperconnected vehicle and is of pressing concern for car manufacturers and related stakeholders.



Manufacturers are making investments in order to equip vehicles with the necessary capabilities and hope to profit from the added value.



Such measures result in an obvious conflict between data ownership and lifecycle management, which is today a challenging debate arena.

Data Quality



Integrity and data quality are a prerequisite to value creation, since applications and services rely on the completeness, accuracy, timeliness, and consistency of data.



One can try to implement reliable and protected sensors and secure and authenticated communication channels, however even hardware may fail or be intentionally maliciously manipulated.



The hyperconnected vehicle owner may directly benefit from falsified data, e.g., in road toll collection, driving behavior-based insurance, or EV charging. Hence, this results to additional conflicts between the observed object and the service providers.

Spontaneous Interaction and Safety

Spontaneous Interaction



Hyperconnected vehicles are exposed to many short-lived interactions, e.g., with other vehicles in order to align routes or signal obstacles.



However, it is near to impossible (or too cumbersome) to authenticate the subjects of such interactions using traditional authentication factors in computer security, since the parties have never communicated before and have no established trust relationship.

Safety



Whereas services are not necessarily a safety-critical system, they are part of the larger ecosystem affecting the vehicle's safety. Integrity is also a fundamental aspect when it comes to safety.



The data collected and processed in the system are utilized by the vehicle for critical decision-making processes.



Clearly, falsified data, being erroneously captured or maliciously modified, endanger the safe operation of the vehicle and the contexts in which it operates, e.g., the traffic. Hence, it is of utmost importance to strive toward guarantees for the integrity of data and by extension of the services that use it.

Computation on encrypted data

Homomorphic encryption allows the service provider to compute the encrypted result of any function and return it to the service provider, but its current performance is disappointing.

Secure computation, e.g., garbled circuits, allow the service provider to compute a predefined function. This is more efficient, but communication intensive, since the communication complexity is on the order of the circuit size that computes the function, and quite inflexible, since the function needs to be predefined.

Property-preserving encryption is very efficient, but somewhat susceptible to leakage-abuse attacks.

Data Perturbation

A common measurement of the degree of perturbation is differential privacy, which provides a guarantee about the influence of an individual value. However, the accuracy of the computation is affected.

Zero-knowledge proofs and verifiable computation

One approach to provide services on data while preserving the integrity is to perform the computation on the client, i.e., decrypt the data and compute any function. In order to prove that the result of the computation matches the encrypted readings, the client provides a zero-knowledge proof or performs a verifiable computation.

Partial observability

In order to validate the sensor readings, one may compare them to other readings, e.g., the ones collected by trusted devices (e.g., roadside units, other vehicles, sensors on the highway, etc.).

However, these public sensors now collect data for every vehicle or other object any time which presents a severe privacy threat.

A special form of authentication allows to strike a balance between these objectives, as it enables ubiquitous surveillance, but only a fraction of the data can be collected. The observed vehicles cannot tell which of their information was collected and which not.

Context-Based Authentication

Vehicles share a common context, e.g., nearby vehicles will have similar acceleration patterns, lighting, and weather conditions, etc. This context can be used to establish an authenticated channel.

A disadvantage of this approach is, of course, that this context is not secret and longer periods of synchronization may be needed.






Mandatory Access Control

Once data are fed back to the vehicle, they are acted upon in the current driving situation.

It is of utmost importance not to endanger the safety of the passengers or their environment (e.g., pedestrians).

Therefore, checks on the actions need to be performed and the checks need to be mandatory and are reminiscent of the access controls put into operating systems.

Application to Use Case

-  V2I map updates carry severe privacy and safety risks.
 -  An incorrect map can significantly increase the physical risks to passengers, e.g., by signaling the need for emergency breaking or simply redirecting traffic into an already congested area.
 -  On the other hand, a centralized infrastructure, e.g., in the cloud creates the opportunity for effective mass surveillance and poses a threat to the passengers' privacy.
 -  Hence, a threat model is followed, where some vehicles may be malicious. In addition, the honest remaining ones are privacy sensitive
-  To implement a real-time V2I map update, the vehicle needs to send their observed events that potentially affect other vehicles to the map service provider. The vehicle's messages need to include at the very least the vehicle's identity, its location, the type, and details of the event. The service provider has to operate on a set of such messages.



Forged identification

The map service provider requires some form of authentication; however, this authentication should not impact the privacy of the passengers.

Hence, a good compromise would be anonymous credentials as already implemented in passports or identity cards.



Forged location

As a countermeasure, the map service provider should require some proof that the provided location is correct. Of course, again this should not negatively impact the privacy of the passengers.

Partial observability can satisfy both requirements. Only a small fraction of locations is revealed, but the vehicles are observed everywhere.

Partial observability requires a penalty for detected misbehavior, and this can be considered in a reputation system that accommodates negative ratings.

Threats by the Update



Forged event

Since the range of possible events and their details is very rich, the falsification may actually be a (non-malicious) error by the sensors. Hence, it is not adequate to secure the sensor hardware and its communication, but also, if possible, assess the quality of data.



Privacy

All the proposed countermeasure against threats by the updating hyperconnected vehicle provided best-effort privacy. However, in the debate about data ownerships, stronger demands can be made and would be in theory technically feasible

Threats by the Map Information



Bulk Download

The obvious countermeasure is to only provide a limited download restricted to the current location and rate limited over time.



Unsafe Map Information

All safety checks need to be mandatory, i.e., complete and impossible to circumvent by the application owner, although the check may implement a waiver, i.e., allowing the system to run despite a failed check.

Threat	Countermeasure	Deployment Challenge
Forged identity	Anonymous credentials	Secure issuance
Forged location	Partial observability	Parameter setting
Forged event	Anonymous reputation system	Parameter setting, updates of cryptographic protocols
Privacy	Encrypted map updates	Currently infeasible due to data amount



Social Debate about Privacy and Parameter Setting

While there is good scientific foundation to choose the key length in encryption, often similar techniques are missing for the choice of parameters in privacy-enhancing technologies.

There are no investigations what would be a good parameter for α in partial observation. When using differential privacy, the choice of ϵ is difficult, while partial progress has been made. Even the choice of k in k -anonymity (for reputation systems) is still difficult.



Updates of Cryptographic Protocols

A rarely discussed drawback of computation on encrypted data is that it is rather difficult to change the protocol.

While inputs to the computation can be easily changed, it does not scale to design a new cryptographic protocol for each update of the algorithm.

Hence, cryptographic protocols need to become as flexible as programs in design and development. A major step toward such flexibility is design of compilers for cryptographic protocols.



Secure Issuance and Key Management

Keys and identities need to be securely issued, revoked, and renewed (lifecycle management).

A prerequisite is to design the software so that keys and associated stored ciphertexts can be easily and securely updated.

This applies to all protocols on encrypted data, but also anonymous credentials for identifying vehicles while preserving privacy.

Conclusion and related works

Related works

Author(date)	Title	Summary of contribution	Implication
Lai C. et al. (2020)	SPIR: A Secure and Privacy-Preserving Incentive Scheme for Reliable Real-Time Map Updates	They proposed a secure and privacy-preserving incentive scheme for reliable real-time map updates to guarantee vehicle users' voluntary participation and real bidding, and make the MSP obtain satisfactory data quantity and quality. In addition, SPIR can achieve anonymity and conditional privacy by using the pseudonym management mechanism.	The authors expanded on previous works to propose an incentive scheme for real time map updates
Makhdoom I. et al. (2020)	PrivySharing: A blockchain-based framework for privacy-preserving and secure data sharing in smart cities	The proposed strategy ensures that personal/critical user data is kept confidential, securely processed and is exposed to the stakeholders on the need to know basis as per user-defined ACL rules embedded in smart contracts. Moreover, the data owners are rewarded for sharing their data with the stakeholders/third parties.	The proposed scheme addresses a number of missing gap in the reviewed paper.

Conclusion



This work discusses upon the hypothesis that it is feasible to ensure integrity, while preserving privacy



In the example use case of multistakeholder interaction, in order to update and make use of dynamic map updates, various approaches that can be taken to strike the balance between “privacy by design” and added value offered by hyperconnected vehicles were presented.

Paper Critique/Discussions

- 1** How will users be incentivised to share the sensing data in this scheme?
- 2** Would a PoC have made the paper better?
- 3** How does the framework relates to/comply with existing fundamental EU GDPR requirements, such as data asset sharing, accessibility and purging with data owner's consent?
- 4** Does the application of blockchain work in this context? Providing a distributed storage of user data without centralized control