Privacy and Integrity Considerations in Hyperconnected Autonomous Vehicles

Stamatis Karnouskos and Florian Kerschbaum

Abstract—**The rapid advances in technology can be witnessed in the emergence of Cyber-Physical Systems that pertain to several domains of our society. In transportation, we see the emergence of self-driving vehicles, that utilize a multitude of sensors and intelligent learning techniques to navigate autonomously. Such vehicles are complex Cyber-Physical Systems that are mobile and due to their sensor and intrinsic intelligence are able to collect, analyze and capitalize upon an unprecedented amount of fine-grained data, as well as collaborate in real-time with multiple stakeholders. Although such rich data can play a key role in data-driven economies of scale, this raises questions with respect to privacy and integrity dependent scenarios. In this work, the feasibility of ensuring integrity, and hence safety, while preserving privacy in the emerging hyper-connected vehicle scenarios is discussed. An exemplary case study on real-time vehicle interactions pertaining to map updates exemplifies the combination of privacy-enhancing technologies with integrity-protecting mechanisms.**

Index Terms—**Cyber-Physical System, Connected Vehicles, Intelligent Transportation System, Security, Privacy, Integrity.**

I. INTRODUCTION

The prevalence of Internet of Things [1] in a multitude of domains, applications and services is proving to be a disruptive one. Never before it is feasible to sense the real world, analyze the data, take informed decisions and act. As the frontiers between physical and cyber world are blurring, innovation can be realized. This is especially evident in the last couple of years in the transportation system domain. Specifically we are seeing the emergence of connected vehicles that feature an impressive array of sensors and on-board decision-making units for increasing their assistance to the drivers [2,3] by providing, e.g., cruise control, parking, collision warning, lane-changing warning, pedestrian detection, platooning and cooperative coordination [4,5].

In addition, the vehicle is increasingly considered a "connected vehicle", that is permanently connected via various communication technologies to the Internet and can also interact with infrastructure via vehicle-to-infrastructure (V2I) services and other vehicles via vehicle-to-vehicle (V2V) services. Typical scenarios of remote diagnostics and vehicle health reporting are now a decade old. A trend towards hyper-connected vehicles, i.e., vehicle-to-everything (V2X) is emerging, where the vehicle beyond V2I and V2V also interacts and exchanges information with any entity capable of doing so, e.g., V2P (vehicle-to-pedestrian), V2D (vehicle-to-device), V2G (vehicle-to-grid) etc. Moreover, we have witnessed the

S. Karnouskos is with SAP, Karlsruhe, Germany. E-mail: stamatis.karnouskos@sap.com.

F. Kerschbaum is with University of Waterloo, Canada. E-mail: florian.kerschbaum@uwaterloo.ca.

application of artificial intelligence, which in conjunction with the sophisticated sensors, leads to the emergence of autonomous self-driving vehicles, and where the driver may be optional in the next decades. In this work, the hyper connected vehicle is in focus, which is well exemplified as 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.

As hyper-connected vehicles at large, and autonomous driving are relative new, security, trust and privacy aspects are not well addressed [6]. Security is often seen as an afterthought, and is more visible due to high-profile attacks with various motivations such as fun, publicity, theft, disruption of operations, etc. However, security and especially privacy and integrity, when involving multi-stakeholder interactions, are still at their dawn. Integrity prevents unauthorized users from modifying or forging data and guarantees that all data is reliable, accurate, consistent and of verifiable quality. Failure to achieve integrity may have severe safety consequences in Cyber-Physical Systems (CPS) as the data and depending services can no longer be trusted or may be maliciously manipulated, which leads to flawed decisions and potential life-endangering actions. 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. Integrity and privacy are pivotal aspects pertaining to key scenarios in hyper-connected autonomous vehicles, but as the complexity increases, securing the components within the vehicle, as well as providing real-world viable solutions for the interaction of the vehicles with third party value-added services is seen as challenging.

This work focuses on this inherent conflict in security objectives. On the one hand, in multi-stakeholder interactions the integrity of the sensed data is key towards ensuring a safe and stable system. Verifying the integrity of sensed data requires checking the possibly remote sensor's reading with contextual readings of other sensors. For example, recently it has been discovered that smart meter readings are often wrong [7]. Comparing a meter's reading to another meter, e.g., an in-network meter, might be a possible countermeasure. However, the straight-forward 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. This led to the "privacy by design" principle.

A recent survey on the public opinion on automated driving reveals that there are worries on safety and privacy aspects

pertaining to hyper-connected vehicles [8]. It may seem one can have only either one – integrity or privacy. This work puts forward the hypothesis that it is feasible to ensure integrity, and hence safety, while preserving privacy in the emerging hyper-connected vehicle scenarios. To this end, it contains a case study on a safety-critical aspect of hyper-connected vehicles, i.e., real-time V2I map updates that focuses on the data and service integrity viewpoints. The scenario and discussions show how to effectively combine privacy-enhancing technologies with integrity-protecting mechanisms.

II. Autonomous Driving and Security

A. The Hyper-connected Vehicle of the Future

The emerging fourth industrial revolution [9] sees in its core Cyber-Physical Systems that pertain several domains and have multi-disciplinary applications. A prominent example of a complex CPS is the autonomous vehicle, that is equipped with a multitude of sensors [6,10] and intelligent logic, that enables it to provide advanced auxiliary services to its users [2,3] and other parties at large. Connectivity is increasingly playing a key role, and it expands towards the realization of hyper-connected vehicles, that interact in real-time not only with their in-vehicle components and services but also with infrastructure, other vehicles and generally any kind of CPS entity (V2X). There are already vehicles on the market such as the Cadillac CTS that feature V2V technology, i.e. Dedicated Short-Range Communications (DSRC) that can handle 1000 messages per second from other connected vehicles in the vicinity of 980 feet [11]. With the prevalence of self-driving vehicles in the next years [12], the hyper-connected vehicle of the future is expected to be a mobile CPS that interacts in a sophisticated manner with its surrounding, and actively participates at multiple levels in processes that pertain to both – the physical as well as the cyber worlds.

The hyper-connected vehicle should be seen as a conglomerate of the high-tech build-in sensors from the manufacturer, as well as external stakeholders. As such a hyper-connected 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., road side units (RSUs), and other vehicles. Hence data generated by the vehicle directly or indirectly as it interacts with other cyber-physical entities and services are gaining importance. Example of such data is travel route, time, speed, environmental conditions measured by the vehicle, travel stops, in-vehicle purchases/payments, changes to route/behavior due to Traffic Message Channel (TMC), V2X communication etc. 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).

The emergence of self-driving autonomous vehicles, is going to further amplify the impact on potential application and services. With more than 10 millions of vehicles with self-driving capabilities by 2020 [13], a paradigm change can be realized. For instance, self-driving electric vehicles can communicate and cooperate among them [14] and with the smart grid, in order to maximize renewable energy utilization and guarantee grid stability. Hyper-connected vehicles fit also well with the IoT smart grid city [15], since, as they roam the city streets, they can provide real-time high-definition measurements on traffic, temperature, CO_2 emissions, air quality, noise, infrastructure inspections etc., all of which can be coupled with appropriate real-time analytics in the cloud, and automatic workflows (e.g., traffic redirection, maintenance etc.) that may lead to better decision-making and continuous optimization of the smart city resources.

The increased number of sensors in hyper-connected cars generate a significant amount of data [6], for instance 25 GB/h [16], that can be assessed and used to take informed decisions. The usefulness of that data may have different life-spans, e.g., used in short-term for navigation, or long-term to determine patterns. There are cases where also short-term data might also be utilized to investigate specific malfunctions etc. However, as it becomes evident, it does not make sense to transport all data (due to quantity as well as usefulness) to a backend, and local processing will need to be applied at the edge (i.e., within the vehicle). As such, the vehicle will evolve to an Edge Data Platform that will complement traditional cloud platforms and services. Access to the data generated by the hyper-connected vehicle, as well as higher-level functionalities of it, may require also a service platform to be available that will enable it to evolve with multi-stakeholder contributions. In addition, that would imply vehicle-specific customizations and behavior, e.g., when interacting with other vehicles. As such, the vehicle is expected to evolve also towards a service platform that would control, creation, orchestration and execution of services. In both cases, the data integrity utilized within the car as well as communicated to other stakeholders is fundamental for a wide range of services and dependent scenarios.

Having the vehicle as an Edge (data & service) platform [17], where services, application and interactions can be realized, would empower a series of futuristic scenarios and potential business models, e.g., remote vehicle diagnostics, cybersecurity, over-the-air system updates, fleet management, and usage-based insurance [18]. Platforms that offer value added services are commercially emerging from the manufacturers such as the Scania One [19]. In addition, with increased intelligence, vehicles will integrate multi-goal objectives extending beyond autonomous driving, e.g., reducing traffic, measure performance of in-vehicle devices, update high-precision maps, measure weather conditions in a smart city, optimize route stops. A key question that emerges is how to reconcile the security goals of multi-stakeholder interactions that result due to the hyper-connectivity of the vehicle, i.e., preserve privacy of the vehicle's passengers and integrity of the infrastructure as whole.

B. Example Scenario: Real-Time Map Updates

There are several scenarios that focus on the security implications for hyper-connected vehicles, and the focus of this work is mostly on higher level services and interactions

among multiple stakeholders in V2X. In autonomous driving scenarios, the self-driving vehicles rely on their embedded software and sensors to navigate successfully [20]. The intrinsic capabilities of the vehicle are also combined with high-precision maps, which can provide additional info on roads and conditions, well before these are detected by the vehicle's own sensors. Furthermore, value-added information, can also be passed to the vehicle ad-hoc from other smart vehicles in the area or infrastructure, e.g., road side units, usually with V2V/V2I communication [21], e.g., when an accident has just occurred on the road.

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, of course, be highly credible, and therefore be verifiable, as any deviation may have devastating effects, e.g., guide the vehicle into an one-way street, effect its driving behavior etc. Such info, is usually static and long-term. In addition, the map may also feature other value-added information, that may be short term (or candidates for long-term which have not yet been verified) such as accidents, temporal traffic jams, slippery conditions, loose material on the road, roadworks, outages of traffic lights, missing traffic signs etc.

The dilemma posed, is how to update the maps and their respective info, with data coming from the field (i.e., the hyper-connected vehicles) which in general cannot be assumed as trusted entities. If such info could be incorporated in a timely manner, the result would be accurate and better quality of maps, that would provide tangible business benefits to multiple stakeholders. Having such active interactions with services that provide high-precision maps, the hyper-connected vehicles can be seen as both consumers of that info, as well as producers, that can play a key role for the benefit of everyone.
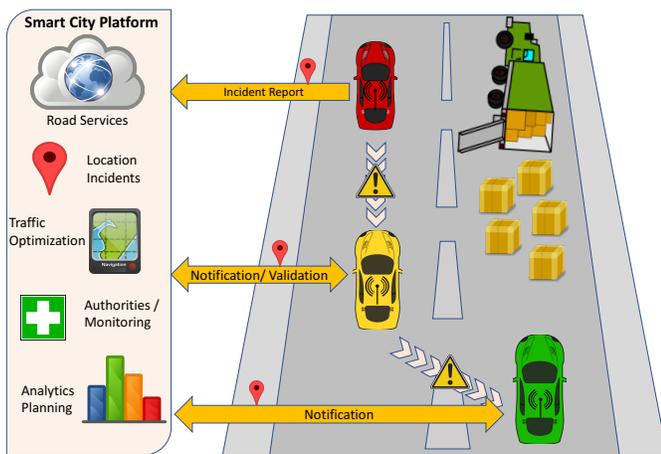

Figure 1. Autonomous vehicles in V2V/V2I scenario

In an example scenario, the in-car maps may reflect a clear two-lane road. However, due to an accident seconds before, some debris from the cargo of a truck are spread on the road, creating a hazardous situation for the passing-by cars as seen

in Figure 1. For the safety of both the car drivers, as well as the people that stop to help in the accident, such info should be communicated asap to the approaching vehicles, which may have no visibility e.g., due to weather conditions, road topology etc. One can argue that there are several ways to do so:

- *In-Vehicle Detection*: Such situational info will be eventually picked up by the hyper-connected vehicle sensor systems, and since it is intelligent, it can recognize the limitations of the obstructed lane for the specific road segment and act accordingly. However, the above consideration, relies heavily on the vehicle's own sensors, and it is not guaranteed that all vehicles will have the same level of sophistication, capabilities, or quick reactions etc.
- *V2V*: A preventive measure would be to propagate such info as quickly as possible via V2V communication, realizing cooperative maneuver planning and cooperative driving [4,5]. In this case, the first vehicle that passes by, assesses the situation, and propagates that info to the next vehicle etc. Eventually this local situation is propagated to all vehicles passing from that point, and therefore they can act in advance, way before their own sensors detect the hazardous situation. However, such scenario depends on the density of the vehicles, and is susceptible to, e.g., misinformation from third parties (other vehicles).
- *V2I*: A complementary measure would be that the vehicle uses its communication facilities (V2I) to inform the respective services that keep high-precision maps on that road segment. Then the maps can be updated, and the info can be pushed to the other hyper-connected vehicles as part of the location specific services, when they are in the vicinity of the affected area. Such scenario enables the information dissemination from a trusted point of distribution, and can also be monitored by remote applications (e.g., monitoring of the highway, city authority control center etc.). Subsequent vehicles that pass by, can verify that the road restrictions are still evident, which would increase the confidence on the validity of the (initially) reported incident. Finally, when new automated reports come in that this limitation no longer exists, the map can again be updated to reflect the new situation and remove the warning.

As it has been discussed, dynamic situations can be detected by vehicle's sensors, and disseminated locally in peer-to-peer ways (V2V) or with infrastructure assistance (V2I), e.g., via cloud based services. The latter is especially valuable for enterprise scenarios, as it enables data collection, analysis and assessment to be carried out at large scale, and has the potential to enhance decision making and planning activities of multiple stakeholders, e.g., city authorities, location service providers, etc. Such dynamic situations can benefit multiple stakeholders in win-win scenarios. However, as it can be easily assumed, they are susceptible to attacks at various levels, where security, trust and privacy play a pivotal role.

Data ownership debates pertaining to hyper-connected vehicles are expected to intensify, as the additional generated data may be controlled by different stakeholders. To what

3

level this will be a market reality is not clear due to the multiple stakeholders involved as, e.g., the vehicle manufacturer collects the aggregated vehicle data, but the subsystem manufacturer (e.g., battery manufacturer) may collect more detailed data pertaining to the usage and performance of the EV battery. The user might also claim data ownership, e.g., because his/her vehicle detected the dynamic situation (such as an accident) and helped by updating the map which may save the lives of passengers of subsequent vehicles. From these considerations, business models can arise, which bring the hyper-connected vehicle to the forefront as both consumer but also producer of valuable information, and pose them as active participants in cross-domain processes, that go beyond its original manufacturer's goals. Especially scenarios such as the map updates discussed here, may have far-reaching enterprise usage, and therefore have the potential to position the future self-driving vehicle in the middle of a data and service based economy. To realize this potential though, the key issues of integrity and privacy need to be sufficiently addressed.

C. Security Objectives and Challenges

The obvious challenges in securing a remotely accessible computer system have already been extensively demonstrated for modern vehicles [10,22]–[25]. However, most countermeasures do not differ in nature from the best practices in secure software development [18,26,27]. Yet, the characteristics of a hyper-connected vehicle – as any system in the future Internet of Things – carry the premise of new, or at least unsolved, challenges in computer security and privacy.

In general, CPS systems [9] are multi-stakeholder, 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. In computer security, the fundamental categorization of confidentiality, integrity and availability (CIA) is followed. Confidentiality refers to the (read) access of data which should only be allowed to authorized users or systems. Integrity refers to the (write) access of data which should ensure that data is correct and up-to-date and can only be modified by authorized users or systems. Availability refers to data or system access being available when and where needed. In hyper-connected vehicles, different distinctive features of these properties can be seen, such as privacy, service offerings, data quality & integrity, spontaneous interactions, and safety.

Privacy in hyper-connected vehicle refers to the passenger's ability to control the use and storage of the data collected by the vehicle. In particular, we 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 not collected data. Privacy is a pressing concern for passengers in hyper-connected vehicles. The hyper-connected vehicle's sensors collect data that allow many sensitive conclusions about its passengers' behavior. The owner of the vehicle therefore has a vested interest and in many jurisdictions the right to control this data. However, even regulations on unobserved lawful access – such as to telecommunication data

in many legislations [28] – are debated. Apart from the direct vehicle users, overall the vehicle is moving towards 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.

Service offering refers to the eco-system of electronic services around a hyper-connected vehicle and is of pressing concern for car manufacturers and related stakeholders. These stakeholders process the hyper-connected vehicle generated data and want to use it to offer new services (or enhance the quality of existing ones), something that undoubtedly improve their business position and stemming benefits. Manufacturers are making investments in order to equip vehicles with the necessary capabilities and hope to profit from the added value. Often however, such measures result in an obvious conflict between data ownership and lifecycle management, which is today a challenging debate arena.

Data quality is an extended form of traditional integrity that also includes problems caused not only by inaccuracies, but also formatting and distribution. 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 & protected sensors and secure & authenticated communication channels, however even hardware may fail or be intentionally maliciously manipulated. In many cases the hyper-connected 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 Interactions are seen as a potential new interaction feature of hyper-connected vehicles. Since these pose an example of mobile CPS, they 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 is one of the most discussed issues pertaining to autonomous cars. Widely accepted standards in automotive industry exist such as ISO 26262 (which is based on IEC 61508) that covers functional safety. With its V framework, it covers (critical) failures and (predictable) hazardous aspects pertaining to product liability. However, when it comes to application of existing standards to autonomous self-driving cars there are several issues flagged, that pose as considerable challenges need to be tackled. For instance, issues raised [29] include the potential non-involvement of the driver, controllability, complex requirements, operational aspects, safety-critical requirements, stochastic system behavior, inductive learning etc. Efforts are underway in several standardization organizations, e.g., via guidebooks for vehicle cybersecurity [30] and hardware protected security [31]. Depending on the level of a hyper-connected vehicle manufacturers can aim at different types of safety. Namely, fail-operational safety is feasible when a driver is still in charge of maneuvering the

vehicle. However, a completely autonomous vehicle should include a fail-safe mode where the vehicle returns to a safe state, e.g., parking road side. It is important that we consider safety not only in the traditional setting of the vehicle, but in the context of the entire eco-system including services. Whereas services are not necessarily a safety-critical system, they are part of the larger eco-system affecting the vehicle's safety. Integrity is also a fundamental aspect when it comes to safety. The data collected and processed in the system is also utilized by the vehicle for critical decision-making processes. For instance, it is used to update the in-vehicle software and services, as well as control other actuators, such as the vehicle lock, brakes and autonomous driving systems. 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 towards guarantees for the integrity of data and by extension of the services that use it.

Generally, on the one hand, CPS is characterized increasingly by a control loop not present at such a large scale in distributed computer systems before. Hence, it is now necessary to protect this control loop against not only random events, but also malicious modifications of data and communication channels. On the other hand, data without control and clear ownership undergo a high risk of abuse. In order to implement subject-verifiable data protection, data minimization is the only reliable principle. Efforts bringing together privacy and integrity ought not to be seen as an operational add-on, but as an integral part of the hyper-connected autonomous car life cycle (from cradle to grave).

III. APPLICABILITY OF SECURITY TECHNIQUES

A. Overview of Techniques

The challenge of securing the future hyper-connected vehicles is to balance these objectives and provide an architecture that combines most of them, meaningfully and realistically. Several security techniques can be used to reconcile the conflicts posed by the objectives; some of which are discussed below, with respect to their advantages or disadvantages, while selected cases exemplify their utilization.

1) Privacy and Service Offering: A fundamental challenge is that the data revealed by a vehicle allows inferences about its passengers and the environment at large. Some inferences may be desired, but not all. Therefore, the challenge is to find a way to allow harvesting the benefits of added value services, and try to prevent unwanted inferences.

a) Computation on Encrypted Data: Obviously, the challenge from revealing data (and meta-data) can be prevented by encrypting the data before sending it to a service provider (or other parties) – while retaining the key. However, in its simplest form this prevents any computation on the data. Homomorphic encryption [32] 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 [33]. Secure computation, e.g., garbled circuits [34], 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 [35] is very efficient, but somewhat susceptible to leakage-abuse attacks [36].

b) Data Perturbation: A more efficient alternative to encryption is the perturbation of data. A common measurement of the degree of perturbation is differential privacy [37], which provides a guarantee about the influence of an individual value. However, the accuracy of the computation is affected. Recently, methods to perturb data at the source have been investigated [38], which would provide a way to protect data emanating from the vehicle. While any function can be computed on the data (and the perturbed result is revealed), accuracy is usually only preserved for few of them.

2) Privacy and Data Integrity: If the data is perturbed or encrypted before sending it out from the vehicle, detecting integrity violations becomes even more difficult, because spotting deviations by calculations at the service provider or manual inspection is no longer feasible. Secure hardware attempts to shield the user from tampering with the system or its data by implementing protection mechanisms in hardware. These protections often are additional casings with seals that trigger alarms when damaged. Nevertheless, even readings that are captured by secure hardware can be corrupted due the secure hardware failing or being circumvented by creating an artificial environment [10], e.g., holding a lighter in the proximity of a fire sensor. Additionally, the storage and communication of data may be tampered with, e.g., destroying a road-toll or an insurance on-board monitoring unit.

a) 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 [39] or performs a verifiable computation [40]. This architecture has been successfully demonstrated for smart meters [41] – an immobile instance of the Internet of Things. Still, the sensor readings themselves need to be integrity-protected, e.g., by secure hardware.

b) 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., road-side units, other vehicles, sensors on the highway etc.). An example would be a vehicle reporting a location, and comparing this location to sensors installed at the road such as a camera. 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 [42], 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. This allows to detect modifications and incentivizes honest behavior (given appropriate penalties).

c) Privacy-Preserving Reputation Systems: A mechanism to deter misbehavior are reputation systems which maintain a score for each vehicle and are based on the

5

hypothesis that past behavior predicts future behavior. The identity of a vehicle can be hidden by anonymous credentials [43]; however, the score (rating) of the vehicle still reveals identifiable information. A countermeasure is to make the score k-anonymous [44]. Of course, protecting the privacy of the vehicles does not prevent the known attacks on reputation systems, such as white-washing and ballot stuffing.

3) Spontaneous Interactions: Context-based Authentication: Hyper-connected vehicles are expected to interact often with each other, and other third-party stakeholders. Such spontaneous interactions occur without the ability to pre-establish trust (or this might be too cumbersome, especially in short-lived interactions). However, 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 [45]. A disadvantage of this approach is, of course, that this context is not secret and longer periods of synchronization may be needed.

4) Safety: Mandatory Access Control: Once data is fed back to the vehicle, it is 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, e.g., no sharp breaking with a trailing vehicle (only if it verified that the trailing is a non-autonomous one that can safely break if timely warned). These checks need to be mandatory and are reminiscent of the access controls put into operating systems. An example is the safe on-board display of information to the driver [46]. One has to consider, that a wide variety of vehicles with varying capabilities, ranging from Level 0 (without any automation), up to Level 5 (full automation) [47]) will be available. As such the heterogeneous mix will feature, e.g., on the one highly autonomous ones (Level 3–5) that can react in a timely fashion, and on the other hand connected ones with some level of automation (Level 1–2) but still human-driven vehicles where the information may be presented to the driver to react.

B. Application to Use Case

In this section a strawman security and privacy architecture is designed, for the use case of a real-time map update (as discussed in subsection II-B). In particular the case of a vehicle-to-infrastructure (V2I) update is considered, where a vehicle driving and observing events is communicating them to a map service provider in the cloud, that then distributes the updates to other hyper-connected vehicles. Of course, these updates can and should be complemented by vehicle-to-vehicle updates in practice.

Such V2I map updates carry severe privacy and safety risks. On the one hand, 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.

In order 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. We place no distinct trust assumptions on the message (except its correct delivery from the vehicle), but instead investigate how different trust assumptions may affect the service providers' and the entire eco-systems operation. The threats surrounding this update message are first considered. It is assumed that a secure network infrastructure for communicating the message is available (as it is typical over the Internet with the TLS protocol) and only threats originating from the end-points of the message are therefore considered.

1) Threats by the update: Potential attacks and proposed countermeasures from the sender of the update, i.e., the vehicle, are first discussed.

a) Forged Identification: The vehicle could provide a forged identity, e.g., pretending to be another existing or contrived vehicle. This could allow the vehicle to send multiple updates in a short time frame. Clearly, 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.

b) Forged Location: The vehicle could provide a forged location of the update, e.g., pretending to be in another city. Even when using secure hardware, the users could, e.g., spoof the GPS signal. This could allow potentially dangerous updates that are difficult to trace back. 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.

c) Forged Event: The vehicle could provide a forged or falsified event in the update, e.g., signaling an accident on a highway where there is none. Clearly, the map service provider needs to verify the event information provided by the vehicles. 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. The later can be realized at the infrastructure side, by checking the information provided against the information by other hyper-connected vehicles and their sensors (it is assumed that major events such as road-blocks, would result in a multitude of reports registering with the respective service).

The map service provider has two options of verifying the updated information – passive and active. In passive verification, the provider waits until it has received sufficiently many updates to make a reliable decision. In an active verification, the provider updates nearby vehicles and asks them to confirm

the information. A combination of both approaches is, of course, feasible, as well.

In order to perform an update of the map and communicate this update to the vehicles, the provider should collect several events. In order to judge the quality of the event a reputation system of vehicles could be used. The reputation system keeps a rating of each vehicle on how credible it was in the past in providing map updates. After a map update has been thoroughly confirmed the map provider sends a positive feedback rating for the vehicle to a reputation provider. The map provider can also query the current rating of the vehicle from the reputation provider. The score to update the map and communicate the changes then should be at least a function of the reputation of the updaters and the safety risk of the event.

The reputation collection should not impact the privacy of the passengers. However, a reputation score may identify a vehicle (think of the most helpful updater). Hence reputation updates should be encrypted [48] and reputation score should be anonymized [44]. Furthermore, a falsified reputation can have severe consequences for the map update, e.g., a single high-reputation vehicle may cause severe harm. Hence, the reputation system should also be verifiable [48], and multiple countermeasures ought to be in place, e.g., not to rely on a single reporting (even if this is a high-reputation vehicle).

d) Privacy: An update not only poses a threat to the map service provider, but also to the entities providing the update. The passengers may be subject to unwanted surveillance. So far, all the proposed countermeasure against threats by the updating hyper-connected vehicle provided best-effort privacy. However, in the debate about data ownerships, stronger demands can be made and would be in theory technically feasible.

In particular, a stronger protection of the location in the update would be to encrypt it – similar to encrypted (anonymous) authentication. However, the location of the update could be inferred from the change in the map and processing the entire map in encrypted form is beyond current computational capabilities. Hence, techniques such as zero-knowledge proof and verifiable computation are not applicable to the map service provider (however they are applicable to the reputation provider). An encryption of the event information currently seems infeasible due to the wide range of possible events and associated information. However, this may change with an increased familiarity of the infrastructure with real-time updates. Alternatives, such as exploiting V2V to obfuscate location tracking, have also been proposed [49].

It should be noted, that all the proposed privacy-preserving techniques follow proper principles of privacy, namely data minimization (revealing only the minimal data necessary), and user verifiability (the vehicle can verify that its privacy is protected).

2) Threats by the map information: Once the map provider has decided to accept an update, the updated map information is communicated back to affected vehicles. This map information poses again several confidentiality and safety risks.

a) Bulk download: An attacker may try to download the entire map of a map provider, e.g., in order to provide similar services himself. The obvious countermeasure is to only provide a limited download restricted to the current location and rate limited over time. The proposed countermeasures of anonymous authentication and partial observability help to ensure that this information is correctly provided and hence the limitations on the download can be reliably and safely implemented.

b) Unsafe map information: Incorrect map information may lead to unsafe driving conditions, which is especially critical if a self-driving, autonomous vehicle relies strongly upon. Hence, it is also necessary for the vehicle to check the consistency of all its information – including the map information, the local sensors and information from other vehicles. No action should be possible that endangers the passengers and depending on the level of autonomy a fail-safe option violating the fail-operational principle may be available (e.g., halting the vehicle on the side of the road). 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. The role of such dynamically updated maps, and how their info is considered by the self-driving vehicle's decision-making processes is seen as challenging, as in critical situations, that would result in life-death decisions to be taken by the vehicle.

IV. Deployment Challenges

While privacy and safety are important objectives for any CPS, deploying a system with safeguards as the ones described in this work faces major obstacles. Privacy cannot be an afterthought in system design, as it is even more complicated to retrofit privacy into an existing design than security. Systems that have been designed with privacy safeguards can always be easily extended with more intrusive functions. Therefore, a best practice is to follow the "privacy by design" principle and built privacy into the design from the start.

Table I
MAP UPDATES: THREATS, COUNTERMEASURES AND CHALLENGES

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

An overview of the threats from map updates by the vehicle is provided in Table I, their proposed countermeasure from subsection III-B and their associated deployment challenges identified in this section. These challenges are elaborated in the following subsections.

A. Social debate about privacy and parameter setting

Privacy is a social good and needs to be balanced between the objectives of service providers and the wider population. Clearly, it is cheaper not to implement privacy safeguards and let the various stakeholders use the data for whatever purpose they desire, including, of course, value-added services.

7

Still, the interest of consumer to maintain or control privacy needs to be respected, and protected. However, not harvesting the benefits of CPS for the consumers and the economy as a whole, is also not a sustainable option. Hence, a social debate needs to take place balancing the conflicting objectives between data use and privacy [50]. Such discussions are not new, and are already ongoing [51,52], but they need to be considered for the context of hyper-connected cars also. This debate will need to achieve a compromise by setting intended parameters for privacy-enhancing technologies.

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. For example, 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, e.g., [53]. Even the choice of k in k-anonymity (as in [44] for reputation systems) is still difficult. Therefore, more research into the implications and proper settings of those parameters is useful and needed to help guide the concurrent social debate.

B. Updates of cryptographic protocols

While an efficient map update is difficult to implement on encrypted data, it is still suggested to process the reputation values on encrypted data. 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 towards such flexibility is design of compilers for cryptographic protocols. There is now a long series of research efforts towards designing such compilers [54]. However, many of these compilers still follow the principle of translating the program into an input for a generic protocol. Specific protocol optimizations are still hard to compile, although also here there is some progress, e.g., [55].

Projecting into the future, assuming the public availability of such compilers there is a need to avoid reducing the privacy guarantees to a trusted third party. If one party designs, programs and compiles the privacy-preserving protocols for hyper-connected vehicles, there needs to be a safeguard in deployment. Clearly, any deployed software with access to the raw sensor data can leak this data to unintended sinks. For example, an independent entity could perform audits of the source code and the build process.

C. Secure issuance and key management

Any security safeguard that relies on cryptography needs to deal with the key management problem. Keys and identities need to be securely issued, revoked and renewed (lifecycle management). This requires a process and authorities to handle the process. 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.

V. CONCLUSION

The hyper-connected vehicles are mobile CPS that can play a pivotal role in several scenarios that go well beyond classical autonomous driving. Due to their sensors and data-management capabilities, they evolve towards edge platforms that can collaborate with a variety of stakeholders, both in their physical vicinity as well as in the cyber plane via V2X interactions. For such real-world utilization, integrity and privacy are key aspects of concern that are raised. This work discusses upon the hypothesis that it is feasible to ensure integrity, while preserving privacy. In the example use case of multi-stakeholder interaction, in order to update and make use of dynamic map updates, it has been shown that there are various approaches that can be taken to strike the balance between "privacy by design" and added-value offered by hyper-connected vehicles. Apart from the technology aspects, it has to be pointed out that there are also social and ethical concerns (which are not seen as in the context of this work) that pertain to the real-world deployment of hyper-connected vehicles, as well as the utilization of their data in the larger contexts of smart cities. Socio-technical aspects need to be adequately discussed and considered, if the visions of autonomous self-driving cars and their expected benefits are to materialize.

REFERENCES

[1] J. Höller, V. Tsiatsis, C. Mulligan, S. Karnouskos, S. Avesand, and D. Boyle, *From Machine-to-Machine to the Internet of Things: Introduction to a New Age of Intelligence*. Elsevier, Apr. 2014.

[2] ETSI, "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions," European Telecommunications Standards Institute (ETSI), Tech. Rep., 2009, ETSI TR 102 638. [Online]. Available: https://goo.gl/4DM77h

[3] NHTSA, "Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application," National Highway Traffic Safety Administration (NHTSA), Tech. Rep., 2014, DOT HS 812 014. [Online]. Available: https://goo.gl/DXh1Nq

[4] M. During and K. Lemmer, "Cooperative maneuver planning for cooperative driving," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 8–22, 2016.

[5] S.-W. Kim and W. Liu, "Cooperative autonomous driving: A mirror neuron inspired intention awareness and cooperative perception approach," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 23–32, 2016.

[6] N. Lu, N. Cheng, N. Zhang, X. Shen, and J. W. Mark, "Connected vehicles: Solutions and challenges," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 289–299, Aug. 2014.

[7] F. Leferink, C. Keyer, and A. Melentjev, "Static energy meter errors caused by conducted electromagnetic interference," *IEEE Electromagnetic Compatibility Magazine*, vol. 5, no. 4, pp. 49–55, 2016.

[8] M. Kyriakidis, R. Happee, and J. de Winter, "Public opinion on automated driving: Results of an international questionnaire among 5000 respondents," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 32, pp. 127–140, Jul. 2015.

[9] A. W. Colombo, S. Karnouskos, O. Kaynak, Y. Shi, and S. Yin, "Industrial Cyberphysical Systems: A Backbone of the Fourth Industrial Revolution," *IEEE Industrial Electronics Magazine*, vol. 11, no. 1, pp. 6–16, Mar. 2017.

[10] J. Petit and S. E. Shladover, "Potential cyberattacks on automated vehicles," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 2, pp. 546–556, Apr. 2015.

[11] Cadillac. (2017) V2V Safety Technology Now Standard on Cadillac CTS Sedans. [Online]. Available: https://goo.gl/b7vkrx

[12] D. Watzenig and M. Horn, Eds., *Automated Driving*. Springer, 2017.

[13] J. Greenough. (2016) 10 million self-driving cars will be on the road by 2020. Business Insider. [Online]. Available: https://goo.gl/e4DbJe

[14] R. Alvaro-Hermana, J. Fraile-Ardanuy, P. J. Zufiria, L. Knapen, and D. Janssens, "Peer to peer energy trading with electric vehicles," *IEEE Intelligent Transportation Systems Magazine*, vol. 8, no. 3, pp. 33–44, 2016.

[15] S. Karnouskos, "The cloud of things empowered smart grid cities," in *Internet of Things based on Smart Objects: Technology, Middleware and Applications*. Springer, 2014, pp. 129–142.

[16] Hitachi, "The Internet on Wheels and Hitachi, Ltd," Hitachi Data Systems, Tech. Rep., 2015. [Online]. Available: https://goo.gl/ofji5u

[17] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[18] PwC, "Connected car report 2016: Opportunities, risk, and turmoil on the road to autonomous vehicles," PwC Strategy&, Tech. Rep., 2016. [Online]. Available: https://goo.gl/TfJmg5

[19] SCANIA. (2017) Scania One introduces connected tool to enhance transport efficiency. [Online]. Available: https://goo.gl/5N3TKb

[20] I. Shim, J. Choi, S. Shin, T.-H. Oh, U. Lee, B. Ahn, D.-G. Choi, D. H. Shim, and I. S. Kweon, "An autonomous driving system for unknown environments using a unified map," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 4, pp. 1999–2013, Aug. 2015.

[21] K. Zheng, Q. Zheng, P. Chatzimisios, W. Xiang, and Y. Zhou, "Heterogeneous vehicular networking: A survey on architecture, challenges, and solutions," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 4, pp. 2377–2396, 2015.

[22] K. Koscher, A. Czeskis, F. Roesner, S. Patel, T. Kohno, S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, and S. Savage, "Experimental security analysis of a modern automobile," in *2010 IEEE Symposium on Security and Privacy*, 2010.

[23] S. Checkoway, D. McCoy, B. Kantor, D. Anderson, H. Shacham, S. Savage, K. Koscher, A. Czeskis, F. Roesner, and T. Kohno, "Comprehensive experimental analyses of automotive attack surfaces," in *Proceedings of the 20th USENIX Conference on Security*, ser. SEC'11, Berkeley, CA, USA, 2011.

[24] A. M. Wyglinski, X. Huang, T. Padir, L. Lai, T. R. Eisenbarth, and K. Venkatasubramanian, "Security of autonomous systems employing embedded computing and sensors," *IEEE Micro*, vol. 33, no. 1, pp. 80–86, Jan. 2013.

[25] M. Amoozadeh, A. Raghuramu, C. nee Chuah, D. Ghosal, H. M. Zhang, J. Rowe, and K. Levitt, "Security vulnerabilities of connected vehicle streams and their impact on cooperative driving," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 126–132, Jun. 2015.

[26] J. Viega and G. McGraw, *Building Secure Software*. Addison Wesley, 2001.

[27] T. Zhang, H. Antunes, and S. Aggarwal, "Defending connected vehicles against malware: Challenges and a solution framework," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 10–21, Feb. 2014.

[28] A. Arnbak. (2013) The politics of the EU court data retention opinion: End to mass surveillance? [Online]. Available: https://goo.gl/kVQikO

[29] P. Koopman and M. Wagner, "Challenges in autonomous vehicle testing and validation," *SAE International Journal of Transportation Safety*, vol. 4, no. 1, pp. 15–24, Apr. 2016.

[30] SAE, "Cybersecurity Guidebook for Cyber-Physical Vehicle Systems," Society of Automotive Engineers (SAE), Tech. Rep., 2016, J3061. [Online]. Available: http://standards.sae.org/wip/j3061/

[31] ——, "Requirements for Hardware-Protected Security for Ground Vehicle Applications," Society of Automotive Engineers (SAE), Tech. Rep., 2015, J3101. [Online]. Available: http://standards.sae.org/wip/j3101/

[32] C. Gentry, "Fully homomorphic encryption using ideal lattices," in *Proceedings of the 41st annual ACM symposium on Symposium on theory of computing - STOC '09*, 2009.

[33] C. Gentry, S. Halevi, and N. P. Smart, "Homomorphic evaluation of the AES circuit," in *Lecture Notes in Computer Science*. Springer, 2012, pp. 850–867.

[34] A. C. Yao, "Protocols for secure computations," in *Proceedings of the 23rd Annual Symposium on Foundations of Computer Science*, ser. SFCS '82. Washington, DC, USA: IEEE Computer Society, 1982, pp. 160–164.

[35] F. Kerschbaum, "Frequency-hiding order-preserving encryption," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security - CCS '15*. ACM, 2015, pp. 656–667.

[36] P. Grubbs, K. Sekniqi, V. Bindschaedler, M. Naveed, and T. Ristenpart, "Leakage-abuse attacks against order-revealing encryption," in *Proceedings of the 38th IEEE Symposium on Security and Privacy (SP)*, May 2017.

[37] C. Dwork, "Differential privacy," in *Automata, Languages and Programming*. Springer, 2006, pp. 1–12.

[38] Z. Qin, Y. Yang, T. Yu, I. Khalil, X. Xiao, and K. Ren, "Heavy hitter estimation over set-valued data with local differential privacy," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*, ser. CCS '16. ACM, 2016, pp. 192–203.

[39] S. Goldwasser, S. Micali, and C. Rackoff, "The knowledge complexity of interactive proof-systems," in *Proceedings of the Seventeenth Annual ACM Symposium on Theory of Computing*, ser. STOC '85. New York, NY, USA: ACM, 1985, pp. 291–304.

[40] R. Gennaro, C. Gentry, and B. Parno, "Non-interactive verifiable computing: Outsourcing computation to untrusted workers," in *Advances in Cryptology – CRYPTO 2010*. Springer, 2010, pp. 465–482.

[41] M. Jawurek, M. Johns, and F. Kerschbaum, "Plug-in privacy for smart metering billing," in *Privacy Enhancing Technologies*. Springer, 2011, pp. 192–210.

[42] F. Kerschbaum and H. W. Lim, "Privacy-preserving observation in public spaces," in *Computer Security – ESORICS 2015*. Springer, 2015, pp. 81–100.

[43] E. Androulaki, S. G. Choi, S. M. Bellovin, and T. Malkin, "Reputation systems for anonymous networks," in *Privacy Enhancing Technologies: 8th International Symposium, PETS 2008 Leuven, Belgium, July 23-25, 2008 Proceedings*. Springer, 2008, pp. 202–218.

[44] S. Clauß, S. Schiffner, and F. Kerschbaum, "k-anonymous reputation," in *Proceedings of the 8th ACM SIGSAC symposium on Information, computer and communications security - ASIA CCS '13*, 2013.

[45] M. Miettinen, N. Asokan, T. D. Nguyen, A.-R. Sadeghi, and M. Sobhani, "Context-based zero-interaction pairing and key evolution for advanced personal devices," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS '14*, 2014.

[46] S. Gansel, S. Schnitzer, A. Gilbeau-Hammoud, V. Friesen, F. Dürr, K. Rothermel, and C. Maihöfer, "An access control concept for novel automotive HMI systems," in *Proceedings of the 19th ACM symposium on Access control models and technologies - SACMAT '14*, 2014.

[47] SAE, "Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles," Society of Automotive Engineers (SAE), Tech. Rep., 2016, J3016. [Online]. Available: http://standards.sae.org/j3016_201609/

[48] F. Kerschbaum, "A verifiable, centralized, coercion-free reputation system," in *Proceedings of the 8th ACM workshop on Privacy in the electronic society - WPES '09*, 2009.

[49] J. Lim, H. Yu, K. Kim, M. Kim, and S.-B. Lee, "Preserving location privacy of connected vehicles with highly accurate location updates," *IEEE Communications Letters*, vol. 21, no. 3, pp. 540–543, Mar. 2017.

[50] D. Anthony, T. Stablein, and E. K. Carian, "Big brother in the information age: Concerns about government information gathering over time," *IEEE Security & Privacy*, vol. 13, no. 4, pp. 12–19, Jul. 2015.

[51] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen, "Security and privacy in smart city applications: Challenges and solutions," *IEEE Communications Magazine*, vol. 55, no. 1, pp. 122–129, Jan. 2017.

[52] M. Musolesi, "Big mobile data mining: Good or evil?" *IEEE Internet Computing*, vol. 18, no. 1, pp. 78–81, Jan. 2014.

[53] J. Lee and C. Clifton, "Differential identifiability," in *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining - KDD '12*, 2012.

[54] D. Malkhi, N. Nisan, B. Pinkas, and Y. Sella, "Fairplay—a secure two-party computation system," in *Proceedings of the 13th Conference on USENIX Security Symposium*, ser. SSYM'04, vol. 13. Berkeley, CA, USA: USENIX Association, 2004.

[55] F. Kerschbaum, "Automatically optimizing secure computation," in *Proceedings of the 18th ACM conference on Computer and communications security - CCS '11*, 2011.

9