Properties of Vehicle Network Privacy

Authors:

1. George Corser, Assistant Professor, Saginaw Valley State University
2. Huirong Fu, Associate Professor, Oakland University
3. Mathias Masasabi, Oakland University
4. Lars Kivari, Oakland University

Contents

Abstract

Contemporary vehicles include a range of safety devices, such as seatbelts, airbags, and even electronic sensors. Future vehicles will also contain wireless devices similar to wi-fi routers which will enable cars and trucks to communicate with each other to avoid crashes. These routers will enable a new type of computer network, the vehicular ad hoc network, or VANET, which may one day save thousands of lives and billions of dollars, reduce fuel consumption and pollution, and expand ubiquitous connectivity and mobile application functionality to the world's roadways. The problem: VANETs may also expose motorists to surveillance by eavesdroppers, from casual stalkers to Big Brother. The problem has confounded researchers for decades perhaps partly because the desired properties of vehicle network privacy have not been sufficiently defined, analyzed and evaluated. The purpose of this paper is to provide a taxonomy to classify privacy properties in vehicular contexts.

1. Introduction and Background

The vehicular ad-hoc network (VANET), sometimes called "The Internet of Cars," presents distinctive privacy challenges. In the United States VANET standards are specified by Dedicated Short Range Communications / Wireless Access in Vehicular Environments (DSRC/WAVE). These standards call for media access control layer (MAC layer) transmissions of precise vehicle locations 10 times per second. DSRC can be used to access the Internet, including LBS applications (APP Layer) which may also require frequent precise location (FPL) data.

Without privacy protections in place, wireless eavesdroppers or malicious LBS administrators could track specific vehicles, or cross-reference vehicles' precise origin and termination geographical coordinates with home and work addresses, using Google Maps or some similar map database, perhaps revealing (*deanonymizing*) a vehicle at a given location at a given time. Because motor vehicles tend to move at high speeds in predictable patterns along prescribed routes, their mobility patterns may make vehicles more vulnerable to location privacy attacks than, say, pedestrian mobile phone users. Deanonymization could occur at either the MAC layer or higher layers. MAC layer VANET systems require vehicles to transmit FPL. LBS applications sometimes have similar requirements. If there is collusion, MAC layer data could be used to circumvent application layer (APP layer) protections such as spatial-temporal cloaking.

The vehicular location privacy problem is important because driver location data can be misused. Employers might monitor an employee's car parked at a competitor's office (revealing an employee's job interview) or at specialized medical facilities (revealing an employee's health condition). It is not difficult to construct further privacy breaches arising from vehicle surveillance by spouses and ex-spouses, or paparazzi and other stalkers. Proposed national level legislation in the United States to address digital location privacy threats includes the Location Privacy Act and the Geolocation Privacy and Surveillance Act.

The location privacy challenge from a technical standpoint is large-scale and complicated in VANETs. Equipment supporting wireless/wifi networks is already being installed in new vehicles. Industry representatives estimate that 90% of vehicles will be wifi-connected within the decade [1]. LBS usage continues to grow rapidly [2] and is expected to expand to VANET platforms [3]. Standards governing VANETs [4] provide data fields for future privacy protocols, but the specifics of these protocols remain open research questions.

Spatial cloaking has been a standard solution to the LBS location tracking problem. The idea is, if k LBS users are operating in a spatial area, s, then k,s-privacy (a derivative form of k-anonymity[26]) is achieved [14]. The problem is, if LBS requests are repeated frequently over time, and only one of the k LBS users is consistent throughout the set of cloaked requests, then that user is exposed. Researchers have modified spatial cloaking to preserve k-anonymity even when LBSs receive frequent requests. However, no research has been performed which addresses the following problems. First, cloaking requires a trusted third party (TTP) or cloaking proxy, which may be unnecessary additional overhead. Cloaking is ineffective in low vehicle densities, especially if only one user is using LBS in the given vicinity.

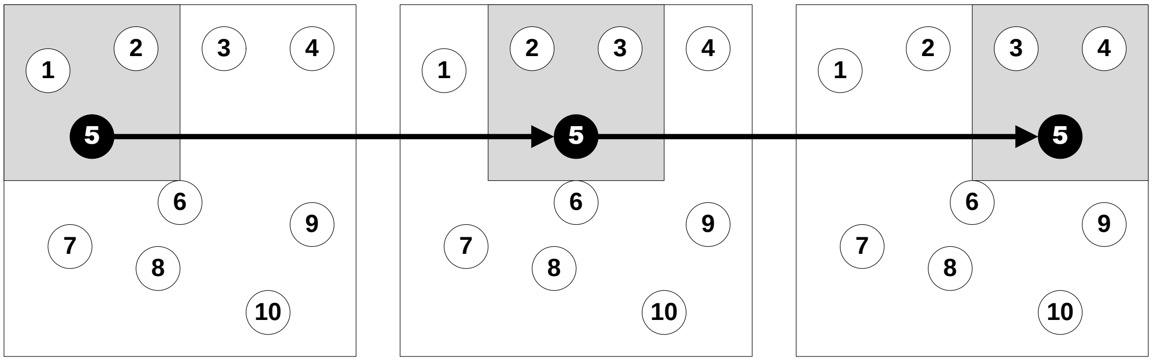


Figure 1: Spatial cloaking in a series of three snapshots: Vehicle 5 maintains *k*,*s*-privacy at each snapshot but all three snapshots analyzed together may reveal the vehicle.

Because of the success of cloaking, other privacy methods remain relatively under-researched. In vehicular settings, *dummy event* and *active decoy* methods may prove especially effective. A dummy event is a message containing false data, sent in order to help conceal a genuine message. Dummy events and genuine messages are sent by the same genuine entity. Dummy events function analogously to aircraft countermeasures, such as flares. An active decoy, on the other hand, is a dummy event sent by an entity other than the genuine one. The proposed research is designed to examine the tradeoffs between safety, efficiency and privacy using dummy event and active decoy methods.

The literature refers to several forms of privacy. *Identity privacy* is sometimes referred to as anonymity, pseudonymity or unlinkability with personally identifiable information (PII). This is often achieved by the use of pseudonyms. *Location privacy* refers to the unlinkability of PII with a geographical position, and further, the unlinkability of one pseudonym with another by using location data. *Query privacy* would make unlinkable to the user's PII, not only the location of the user, but also the particular query made or query service used. Privacy must be constrainable, as in the cases of *conditional privacy* and *revocability*. This paper focuses on identity privacy and location privacy.

The rest of this paper is organized as follows. Section 2, Privacy Properties, discusses the theoretical underpinnings of vehicular location privacy and describes many of its desired properties and existing solutions. Section 3, Future Directions, suggests new ares for vehicle privacy research. Section 4, Conclusion, concludes the paper.

2. Vehicle Network Privacy Properties

Privacy concerns are subordinate to safety considerations. Vehicle safety has historically been a matter of *crash mitigation*. Safety belts, air bags, collapsible steering wheels and other technologies have been designed to reduce the severity of the consequences of a crash. As the figure below suggests, crash mitigation seems to have reached a plateau.

A new direction in automotive safety has arisen in *crash prevention*. Vehicle networks have been proposed which would allow every car to compute its trajectory and the trajectories of other vehicles to alert drivers regarding potential crashes faster than human response alone would achieve. The US DOT reports such technology could eliminate 80% of crashes of unimpaired motorists [30]. Besides saving lives, crash prevention technologies such as those predicted in vehicle networks, if effective, may reduce the price of cars. Expensive crash mitigation components, like airbags, may become unnecessary, and may be superseded by more effective crash avoidance components.

> *"It will change driving as we know it over time," said Scott Belcher, president and CEO of the Intelligent Transportation Society of America. "Over time, we'll see a reduction in crashes. Automobile makers will rethink how they design and construct cars because they will no longer be constructing cars to survive a crash, but building them to avoid a crash." [31]*

To achieve network latencies far faster traditional ones, a new set of protocols were developed and a new spectrum assigned specifically for vehicles. The Federal Communications Commission (FCC) dedicated a 75 MHz spectrum in the 5.9 GHz band for DSRC.

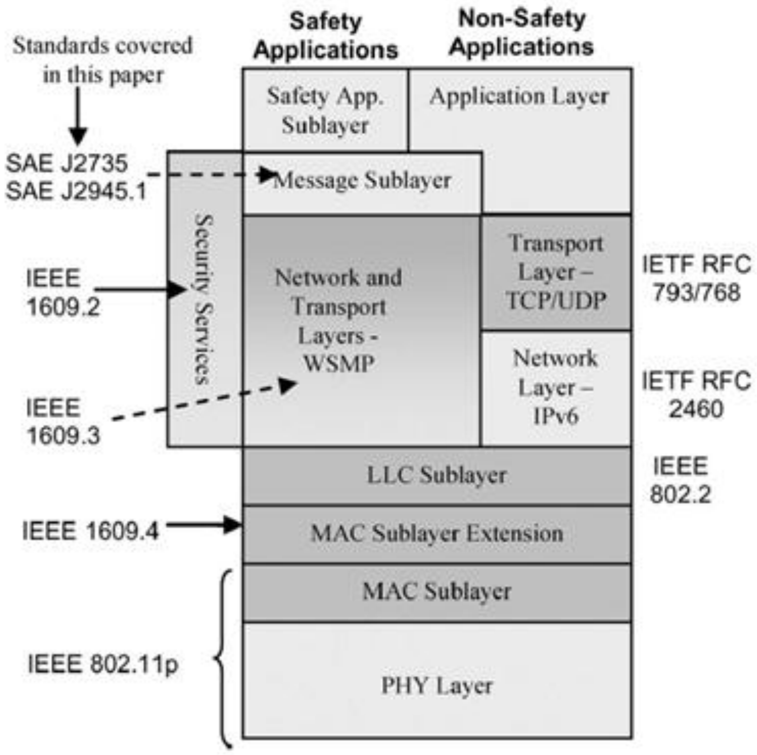


Figure 2: DSRC protocol stacks (Kenney, 2011)

DSRC features two distinct network/transport layer protocols. The first is called WAVE, Wireless Access for Vehicular Environments, which features WSMP, WAVE Short Message Protocol, which would typically be used in V2V safety applications. The second, IPv6/TCP/UDP typically would be used in V2I, especially when accessing infotainment or enables safety services which use SAE J2735 message protocols that include a message type called a Basic Safety Message, BSM, also referred to as a heartbeat message.

Property 1: Collision Avoidance. BSMs must maximize safety.

To achieve collision avoidance, vehicles inform each other of whether or not they are on a trajectory to collide. In VANETs this is accomplished using BSMs. The BSM is the fundamental building block of VANET safety systems, and the fundamental privacy vulnerability addressed by the privacy protocols evaluated in this research. In the United States, the Society of Automotive Engineers (SAE) has established a standard, J2735, the DSRC Message Set Dictionary, which specifies the message sets and data elements for VANET communications. SAE J2945 specifies the minimum performance requirements; SAE J2945-1 specifies the requirements for the BSM [32]. This exposition uses only the American terminology, though similar standards are set by European Telecommunications Standards Institute (ETSI) and the European Committee for Standardization (CEN), and by the Japanese Association of Radio Industries and Businesses (ARIB).

BSMs must be fast, frequent, and localized, and include the transmitting vehicle's precise position, velocity and direction at precise times. For BSMs, *fast* means ultra low-latency wireless communication using 802.11p which omits BSS and associated MAC sublayer overhead [5]. An average-sized 320-byte message sent over a standard 6 Mbps channel would take 4.27 ms to transmit; no handshaking or acknowledgement is required. *Frequent* means each vehicle transmits every 100 to 300 ms, depending on localized range and vehicle density. *Localized* means a vehicle transmits within a 15 m to 300 m radius [33], depending upon vehicle speed and the number of other vehicles in the vicinity, which may contribute to wireless network congestion. *Precise position* is measured by the error range of the location reported by a vehicle's global positioning system (GPS).

Velocity and *direction* can be computed from multiple precise positions. One manufacturer [34] of on-board equipment (OBE) reports positional accuracy of 2.5 m, CEP50, i.e. 50% of measurements are within the circular error probability (CEP) range of 2.5 m. Vehicular positional accuracy continues to be an active area of research. Methods have been proposed to sharpen the accuracy of GPS using ancillary systems, such as RFID-assisted localization systems [35]. Since GPS can suffer from interruption or interference, non-GPS solutions [36] have also been suggested. IEEE 1609.4 specifies GPS as an effective source of *precise time*. The preciseness and frequency of BSMs are absolutely required for safety applications, but also represent fundamental vulnerabilities to privacy.

Privacy protocols diminish safety when they include a *silent period*, i.e. a time span when no BSMs are transmitted. The importance of availability of the safety system can be illustrated using rough figures from government and NGO reports. According to the NHTSA Fatality Analysis Reporting System [37] in 2011 there were 1.10 fatalities per 100 million vehicle miles traveled. There were 3 trillion miles traveled according to the Federal Highway Administration (FHWA) [38], which equates to roughly 33,000 fatalities. If we estimate 82% of fatalities could be eliminated using BSMs without silent periods [5], then the system would save 27,060 lives. If a privacy protocol required all vehicles to activate a BSM silent period 10% of the time, then the cost to safety would be 2,706 lives per year. These are a naïve estimates, of course, but they serve to illustrate the impact of unavailability of BSMs. We hope more accurate estimates will be published in future research.

Property 2: Authentication. BSMs must be trustworthy.

Vehicles must be able to rely on each others' BSMs. It is assumed that non-malicious vehicles will attempt to transmit accurate BSMs because each vehicle's safety depends on its neighboring vehicles knowing its precise position, and vice-versa. However, inaccurate BSMs may be transmitted, either accidentally or maliciously. It is possible that a vehicle's damaged or defective GPS could accidentally generate inaccurate location readings. Cryptographic authentication would not solve this problem because the messages would be authentic. Authentication is designed to protect against attackers who may maliciously transmit false BSMs, degrading traffic flow or perhaps even inducing vehicle collisions.

To achieve authentication, BSMs must include valid *digital signatures*. Because digital signatures provide quick, low-overhead authentication, they have long been accepted as the most effective mechanism to ensure authentication and message integrity in vehicular environments where nodes may often be in transmission range of each other for only a brief time [33]. IEEE 1609.2 specifies data structures and other standards for authenticating BSMs, such as Elliptic Curve Digital Signature Algorithms, ECDSA-224/256 and SHA-256 hash algorithm. IEEE 1609.2 does not describe the broader PKI system. Often the broader PKI is discussed in the privacy protocol proposals.

Digital signatures do not provide confidentiality, but confidentiality is not important. BSM data do not contain secret information. In fact, BSM data are designed to be transparent for safety reasons, so confidentiality would be counterproductive. However, this lack of confidentiality, coupled with verification of credentials provided by the digital signature, also makes BSMs vulnerable to privacy attacks.

Privacy protocols diminish authentication effectiveness when they introduce vulnerabilities or impair performance in processing or accessing digital signatures.

Some proposals discuss a tradeoff between storage, computation time and transmission time of digital signatures. In our evaluation we do not consider dollar costs of private key storage. We only consider availability, confidentiality and integrity of the keys and digital certificates. To protect against attackers who might attempt to steal private keys used for digital signatures, some have proposed that such secret information be stored in a *tamper proof device* (TPD) in each vehicle. Such a component in a privacy system would tend to bolster its rating. To protect against both accidental and malicious BSMs it has been proposed that neighboring vehicles' BSM data be confirmed by other sensors, such as short range radar or cameras similar to those used by self-parking cars, or that neighboring vehicles be measured for their trustworthiness by reputation-based trust systems [33]. This and any other method which explicitly improves the trustworthiness of digital signatures would receive a commensurately improved authentication rating.

By our definition, the property of *authentication* is separate from the properties of *accountability* and *revocability*, described below. A privacy protocol which requires only one single CA, for example, may centralize key distribution which may simplify certificate revocation (revocability) or may help law enforcers identify culpable vehicles involved in traffic accidents (accountability), but it may also introduce a single point of attack and a performance bottleneck, a threat to the availability and integrity of the digital signatures (authentication). In such a case, we would not classify the privacy protocol as high-trust though we might rate it highly in other respects.

Digital signatures defined in IEEE 1609.2 include Provider Service Identifier (PSID). The PSID indicates the application that will use the data in the payload, so the PSID is analogous to a port number in TCP. It is possible not all BSMs will contain identical PSIDs, in which case an eavesdropper might be able to track a vehicle by using PSIDs within the digital signature. All DSRC privacy protocols suffer from this particular deficiency, so we do not consider it when determining our rating.

Property 3: Pseudonymity. BSMs must not reveal real identities of vehicles or owners.

To achieve pseudonymity, a type of *identity privacy*, BSMs must use pseudonyms, or *pseudoIDs*, each of which having a corresponding digital certificate. Except in circumstances requiring *accountability* or *revocability*, described below, pseudoIDs and their certificates are *unlinkable* to the vehicle identification number (VIN) of the vehicle and to the personally identifiable information (PII) of the vehicle owner.

In the literature the term, *unlinkability*, may refer to the inability to correlate vehicle identities with pseudoIDs, but it may also refer to the inability to correlate between multiple pseudoIDs of a particular vehicle. To avoid ambiguity, we refer to the former as *pseudonymity* and to the latter as either *untrackability* or *untraceability*, defined below.

Privacy protocols diminish pseudonymity when they risk linkage between pseudoID and VIN or PII. There is a natural tradeoff between authentication by digital signature and pseudonymity by pseudoID. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for pseudonymity, the more the better.

Property 4: Untrackability. PseudoIDs in currently transmitted BSMs must not be linkable to pseudoIDs in immediately preceding BSMs from the same vehicle.

If a vehicle were identified (*marked*), and its pseudoID linked to PII even a single time, then the vehicle could be tracked as long as its BSM used that same pseudoID.

To achieve *untrackability*, a type of *location privacy*, BSMs must use multiple pseudoIDs, rotating between them frequently, on average every 5-10 minutes. A single vehicle may contain several, or even thousands of pseudoIDs, each with its own digital certificate. By periodically changing between many pseudoIDs theoretically a vehicle could only be tracked while a particular pseudoID was in use subsequent to the vehicle being marked [5].

Privacy protocols diminish untrackability when they risk linkage between current pseudoIDs and their immediately preceding pseudoIDs. There is a natural tradeoff between authentication by digital signature and untrackability. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for untrackability, the more the better. If a pseudoID switching technique in a privacy protocol includes a silent period when no BSMs are transmitted, there could also be a tradeoff between collision avoidance and untrackability.

Property 5: Untraceability. PseudoIDs in current or past BSMs must not be linkable to other pseudoIDs from the same vehicle, except by proper authorities.

To achieve *untraceability*, another type of *location privacy*, sometimes called *historical location privacy*, BSMs must use multiple pseudoIDs, switching between them, as in untrackability, above. However, the property of untraceability is distinct from untrackability. By our definitions, tracking a vehicle would be performed in real-time, while the vehicle is in motion. Tracing the vehicle would be a matter of historical investigation, to determine what vehicle was at what location at what time. This sort of evidence-gathering has been used by proper authorities, such as courts of law (see *accountability*, below).

But tracing could also be used by stalkers or paparazzi for gathering background information on vehicles to establish locations at specific times or to establish transportation patterns of people under unauthorized surveillance. Sometimes in the literature definitions of the terms untrackability and untraceability are interchanged. Sometimes they are used as synonyms. The properties are similar but not exactly the same. We offer our definitions as standard terminology to distinguish between protocols protecting real-time location privacy (untrackability) and historical location privacy (untraceability).

Privacy protocols diminish untraceability when they risk linkage between pseudoIDs and preceding pseudoIDs for a given vehicle. There is a natural tradeoff between authentication by digital signature and untraceability. For authentication to be fast and efficient, the fewer the pseudoIDs and certificates the better; for untraceability, the more the better. If a pseudoID switching technique in a privacy protocol includes a silent period when no BSMs are transmitted, there could also be a tradeoff between collision avoidance and untraceability.

Property 6: Accountability. PseudoIDs must be linkable to PII by proper authorities.

Sometimes it is beneficial to link a vehicle to its owner's identity and/or its location, such as when a vehicle may have been used in a crime or involved in an accident. It may be argued that a privacy protocol without the property of accountability would introduce more risk to the public by concealing criminals than it would introduce security to the public by protecting people's privacy.

To achieve accountability, a certificate authority (CA) or other trusted third party (TTP) must protect vehicle and owner identity and location while maintaining the capability to link this information with pseudoIDs if requested by a proper authority. This is sometimes referred to as *conditional privacy*.

Privacy protocols diminish accountability when they do not provide a secure mechanism for linkage between pseudoIDs and vehicle/owner identity and location. There is a natural tradeoff between trust and privacy, and this balance must be struck in the property of accountability. The TTP must be able to determine the circumstances under which a proper authority may circumvent a privacy protocol and reveal the true identity associated with a pseudoID.

Property 7: Revocability. PseudoIDs and digital certificates must be rescindable.

It is possible that valid digital certificates could be stolen and used maliciously. If this is detected the certificate should be revoked.

To achieve revocability, a CA or other TTP must provide valid digital certificates for pseudoIDs while maintaining the capability of rescinding certificates by updating and distributing a *certificate revocation list* (CRL) if requested by a proper authority.

Privacy protocols diminish revocability when they impair the distribution of CRLs securely, quickly and broadly [39]. For authentication to be fast and efficient, the smaller the CRLs, the better; for effective revocability, some protocols indicate large CRLs. There is a natural tradeoff between trust and privacy, and this balance must be struck in the property of revocability. The TTP must be able to determine the circumstances under which more harm than good comes from BSMs bearing a particular pseudoID and that the benefit of revoking that pseudoID's digital certificate exceeds its cost.

Property 8: Anonymity. Privacy models must maximize indistinguishability between pseudoIDs.

Privacy protocols can be evaluated by *anonymity*, which we define as the quantifiable amount of privacy the vehicle's pseudoID enjoys by using the protocol. Anonymity could measure identity privacy or location privacy. The pseudoID is the mechanism which protects identity privacy, therefore it follows that pseudonym anonymity could measure identity privacy protection. But what about location privacy?

Shokri enumerates four methods of preserving location privacy: obfuscation, hiding events, adding dummy events, and anonymization [10]. For BSMs, obfuscation is not possible because precision is required for safety applications. Hiding BSMs is not possible because safety applications depend on the detectability of BSMs. Adding dummy BSMs may threaten safety by inducing vehicles to react to nonexistent vehicles; in fact digital signatures are used to reduce the possibility of malicious fake BSMs. The only remaining method is anonymization. Since the identities used in BSM transmissions are pseudonyms, pseudoIDs, the only way to protect privacy in VANETs is by *pseudonym anonymity*, or as we call it, anonymity.

It is necessary to make fine distinctions between the terms, anonymous and pseudonymous, to clarify computational privacy in vehicular network contexts. The dictionary definition of *anonymous* is, "not named or identified" [40], a definition which cannot apply in vehicular networks that require identifiers. When referring to computers the term, anonymous, sometimes means using a pseudonym which is unlinkable to a person's true identity, as in an *anonymous post* on a blog. This definition introduces ambiguity with the term anonymity when used as in *anonymity set*, defined below. We use the dictionary definition of *pseudonymous*, "bearing or using a fictitious name" [41] to indicate unlinkability to PII. Vehicle networks use pseudoIDs, which achieve the property of *pseudonymity*. In this paper we use the term anonymity as it is used in set theory, as in an anonymity set [42]. Thus we can define two distinct privacy properties, pseudonymity and pseudonym anonymity.

To achieve anonymity, privacy models must maximize the *anonymity set size* of each pseudoID.

An anonymity set (AS) is a group of entities indistinguishable from one another. Anonymity set size, $|AS|$, is the number of entities in that group. To achieve pseudonym anonymity, privacy models must create conditions where $|AS| > 1$ for BSM pseudoIDs. This is challenging because of the rapid repetition of BSMs containing the same pseudoID, and because BSMs contain data fields which may uniquely or partially identify a vehicle.

Incidentally, one application of the anonymity set concept is *k-anonymity*, which requires that in the results of a database query each entity must be indistinguishable from $k - 1$ other entities [26]. Gruteser and Grunwald apply *k*-anonymity to vehicular location privacy [27], but this too is in the context of database queries. Some researchers have challenged the appropriateness of *k*-anonymity, which depends on a centralized anonymity server (CAS) to obfuscate queries, as a valid metric for location privacy [43]. In the context of BSMs, under our threat model, there is no way to enforce *k*-anonymity. If an antenna were set up at a particular intersection, it could be used to record and log the BSMs of vehicles. The logs of multiple antennae could be used to track or trace any or all vehicles in their vicinities by following the pseudoIDs contained in BSMs. Whoever sets up the antennae would control the logs, and anyone could set up antennae. No one could enforce obfuscation in queries.

BSMs with the same pseudoID are transmitted frequently, every 100 ms, which at 60 mph (100 kph) is every 8.8 ft per 100 ms (2.8 m per 100 ms). A Ford F-150 pickup truck is about 18 feet (5.5 m) long, so even at relatively high speeds, if one were to take top view snapshots of an F-150 the instant it transmitted BSMs and superimpose the images on a map, the snapshots would overlap even if the vehicle were traveling at highway speeds. Rapid repetition of BSMs simplifies tracking for attackers because even if all vehicles in an area were suddenly to change their pseudoIDs an eavesdropper could tell which preceding pseudoIDs correlated to their successors by comparing the nearest most likely previous positions.

Privacy protocols address the BSM tracking and tracing problems by periodically changing pseudoIDs, as discussed above. Privacy protocols address the BSM pseudoID anonymity problem using protocol-specific pseudonym-changing techniques, usually involving mix zones, silent periods and/or group signatures.

Mix zones are geographical areas where vehicles cannot be detected. In mix zones multiple vehicles are in close proximity. Vehicles may or may not change the direction of their trajectories, but they will change their pseudoIDs [20]. Due to the precision of BSM location data, mix zones without silent periods provide little or no effectiveness in achieving anonymity.

Silent periods, as mentioned above, are time spans when no BSMs are transmitted and when vehicles change pseudoIDs. A silent period without a mix zone (area of multiple vehicles) is of minimal use, because if there is only one vehicle in an area, changing pseudoIDs will not fool anyone.

Group signatures are authentications of BSMs made by a key shared by multiple vehicles. In the *group model* vehicles travel in clusters, all using the same group temporary identifier, and authenticating messages using the same group signature [8]. In simulations the group model has been shown to be effective in achieving anonymity in VANET communications but it is less effective the lower the vehicle density, since anonymity level depends on the number of vehicles in each group. The group model introduces inefficiency by additional overhead for group setup and group join/verify processes. Some researchers suggest the group model is infeasible due to limitations of bandwidth and computation power, since pseudoID schemes create large *certificate revocation lists*, CRLs, and associated *checking costs*, network overhead necessary to verify that certificates have not been revoked [9].

Silent periods may impair the safety property. Group signatures may impair digital signature efficiency, the trust property. The method used by the protocol to achieve anonymity was not a factor in our evaluation of a protocol's anonymity property, though it may have affected our evaluations of other properties.

Anonymity has been measured using a range of metrics, not all of which apply to our evaluation. Because we assume no central capability to obfuscate data, a wide range of database query metrics do not apply, including k-anonymity [26], l-diversity [17], t-closeness [18], L1 similarity [44], m-invariance [45][46], and ε-differential privacy [19]. We also set aside network metrics such as combinatorial anonymity degree (CAD), zone-based receiver k-anonymity (ZRK) and evidence theory anonymity [47]. These are all valuable measures of anonymity but they do not measure pseudonym anonymity in the context of BSMs. We evaluate pseudoID anonymity of privacy protocols using one or more of the following metrics: pseudonym anonymity set size, *|AS|;* entropy of the anonymity set size, *H(|AS|),* also called entropy anonymity degree (EAD) [47]; and tracking probability, *Pt*. Based on the best evidence available we evaluated each privacy protocol on its ability to achieve indistinguishability amongst pseudoIDs. We used the following definitions.

Anonymity set size. The anonymity set, AS_i, of target LBS user, *i*, is the collection of all LBS users, *j*, including *i*, within the set of all LBS userIDs, *ID*, whose trajectories, T_j , are indistinguishable from T_i .

$$AS_i = \{ j \mid j \in ID,\ \exists T_j \ s.t.\ \ p(i,j) \neq 0\} \tag{1}$$

Entropy of the anonymity set size. Entropy represents the level of uncertainty in the correlations between trajectory T_i and trajectories T_j. The entropy H_i of the anonymity set AS_i is:

$$H_i = -\sum_{j \in AS_i} p(i,j) \times \log_2(p(i,j)) \tag{2}$$

Tracking probability. Tracking probability, Pt_i, is the probability that the size of the anonymity set of a vehicle under surveillance is equal to one, which can be written as follows.

$$Pt_i = P(|AS_i| = 1) \tag{3}$$

The tracking probability metric is important because average *Pt* tells what percentage of vehicles have privacy, not just how much privacy exists in the overall system. If AS_i = 1.0, then vehicle *i* has no privacy, no defense against tracking.

Privacy protocols diminish anonymity when they do not provide high levels of indistinguishability between pseudoIDs, as measured by *|AS|*, *H(|AS|)* and *Pt.*

Just as digital signatures may introduce privacy vulnerabilities, as discussed in authentication, above, BSMs may introduce privacy vulnerabilities which could be used to reduce anonymity set size by isolating characteristics peculiar to a specific target vehicle. Some of the data fields in a BSM, for example, could be used by eavesdroppers to link past pseudoIDs to current ones. If `MsgCount`, a sequence number field used for packet error monitoring, is not reset when pseudoID changes then an eavesdropper could match the sequence numbers of past BSMs with current BSMs and link the pseudoIDs. If `PositionalAccuracy` differed from vehicle to vehicle, then an eavesdropper could link pseudoIDs or even track this data element instead of tracking pseudoIDs. If

`SteeringWheelAngle` were misaligned for a vehicle, or if `BrakeSystemStatus` were in any way distinct for one vehicle compared to its neighbors it make an easy mark. Perhaps the most concerning data element is `VehicleSize`, which provides vehicle length and width accurate to a resolution of 1 cm. Vehicles are produced in sufficiently various lengths and widths that this element alone may be sufficient for an eavesdropper to link past and present pseudoIDs. All DSRC privacy protocols suffer from this particular deficiency, so we do not consider it when determining our rating, but the challenge remains: how to fine-grained safety data without exposing driver location.

The property of anonymity can be measured at both the APP layer and the MAC layer by the equations in the prior section and the equations below. Since the former were introduced in MAC layer literature, they are included above. The latter were introduced in an APP layer publication [15], so they are included below.

Short term disclosure (*SD*) is a measure of the probability of an eavesdropper successfully identifying any particular true location given a set of true and dummy locations over a presumably short time. If there are *m* time slots and D_i is the set of true and dummy locations at time slot *i*, where $| D_i |$ is the size of D_i, then

$$SD = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{|Di|} \tag{4}$$

Long term disclosure (*LD*) is a measure of the probability of an eavesdropper successfully identifying a true trajectory given a set of true and dummy trajectories. The more trajectories overlap, the lower the probability of detection. If there are *n* total trajectories and *k* trajectories that overlap, then there are $n – k$ trajectories that do not overlap. If T_k is the number of possible trajectories amongst the overlapping trajectories, then

$$LD = 1 / (T_k + (n – k)) \tag{5}$$

Distance deviation (*dst*) is the average of distance between trajectories of dummies and the true user. To define dst_i as the distance deviation of user *i*, let PL^j_i be the location of user *i* at the *j*th time slot and let L^j_{dk} be the location of the *k*th dummy at the *j*th time slot. The function *dist()* express the distance between the true user location and the dummy locations. Then

$$dst_i = \frac{1}{m} * \frac{1}{n} * \sum_{k=1}^{n} \sum_{j=1}^{m} dist(PL^j_{i,}\ L^j_{dk}) \tag{6}$$

Property 9: Decentralization. BSMs must not be traceable by a single TTP.

To achieve decentralization, BSMs must include *blind signatures*, which require a traceable anonymous certificate (TAC). A TAC is a certificate issued by two TTPs, a Blind Issuer (BI) and an Anonymity Issuer (AI). Therefore, decentralized protocols by our definition must call for multiple independent TTPs. As far as authentication is concerned the certificate works the same as any digital signature, but to trace back to the PII of the vehicle requires the cooperation of multiple TTPs. For a full discussion, see [48] and [49].

The purpose of decentralization is to prevent a single TTP from being able to trace or track vehicles. Several researchers have referred to this as Big Brother, so we include in this property of decentralization the concept of Big Brother defensibility. Technically, if both the BI and the AI were controlled by the same administrative authority, it would be possible to have decentralization without Big Brother defensibility.

Privacy protocols diminish decentralization when they fail to call for multiple independent TTPs.

Property 10. Map Database Undeanonymizability. Privacy models must minimize map database undeanonymizability.

LBS applications such as Google Maps API or US Census TIGER database may be used to convert longitude and latitude coordinates into postal addresses, that is, if a vehicle attempted to send fake coordinates to an LBS, the LBS could to *deanonymize* the data. Protected data must be, to the extent possible, un-deanonymizable. Undeanonymizability may require application level defenses, such as EPZ (endpoint protection zones) [28].

3. Future Directions

Shokri [10] suggests there are four methods for preserving location privacy: hiding events, adding dummy events, obfuscation and anonymization.

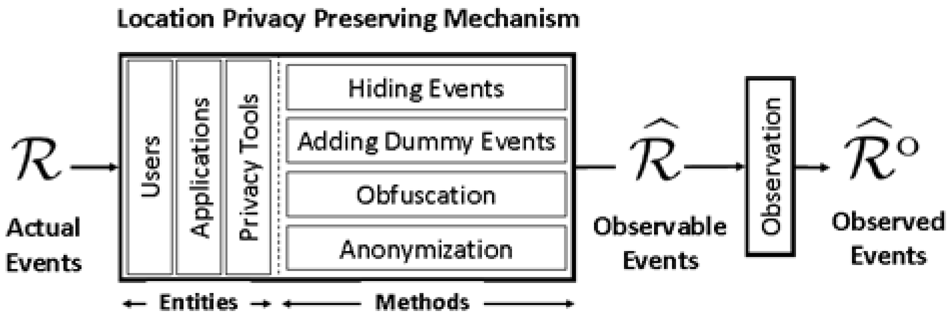


Figure 3: Location privacy preserving mechanisms [10]

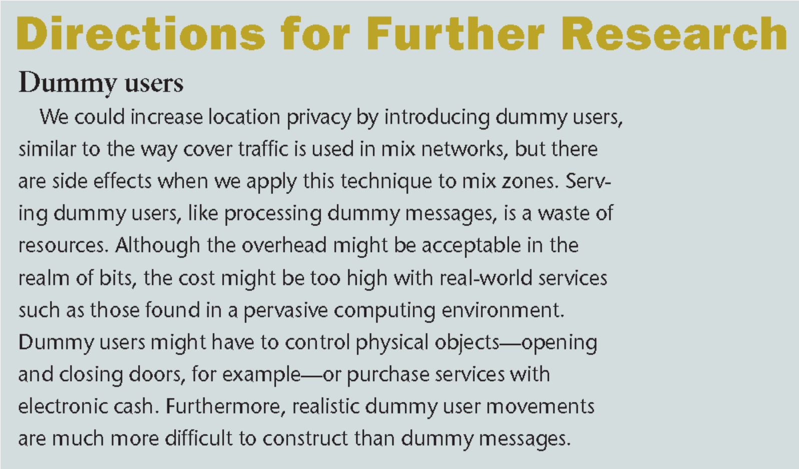
Methods can be used individually or in combination to develop defenses. Methods vary in effectiveness in vehicular settings at the MAC and APP layers, as shown in Table 1.

All techniques listed in the table, except active decoy, impair APP layer continuous precise location (CPL) and frequent precise location (FPL) queries. Other problems include anonymizing deficiencies (PseudoID-to-pseudoID tracking, map deanonymization) and MAC layer cloaking/decoy problems (too slow for safety beacon, exposes duplicate beacons, complicates authentication/CRL/congestion).

Table 1: Location privacy, vehicular methods and techniques

Method	Technique	MAC Layer	IP, APP Layers
Hiding	Silent period	Unsafe	No service
+ Anonymizing	Mix zone	Unsafe	No service
Anonymizing	PseudoID	OK	OK
+ Dummifying	Cloaking region	Latency:TTP	Congestion
+ Dummifying	Active decoy	OK	Congestion
Dummifying	False data	Unsafe	Congestion
Obfuscating	Noise	Unsafe	Impaired service

The literature relevant to vehicular location privacy research can generally be divided into two areas: papers which present MAC layer solutions to narrow vehicle-network-specific technical challenges, and papers which present broadly applicable theoretical concepts of location privacy. The vehicle-network-specific literature largely ignores dummy event solutions for aforementioned safety reasons. The broader theoretical literature ignores dummy event solutions at least partly because of the Privacy Universality assumption, the idea that all or most users desire privacy. The quote below, for example, appears to make this assumption in pervasive computing environments. This may not be the case. If privacy is desired by only a small subset of people, and not by all or many people, then it may be erroneous to conclude that dummy events would cause excessive overhead, even in pervasive computing environments.



Directions for Further Research

Dummy users

We could increase location privacy by introducing dummy users, similar to the way cover traffic is used in mix networks, but there are side effects when we apply this technique to mix zones. Serving dummy users, like processing dummy messages, is a waste of resources. Although the overhead might be acceptable in the realm of bits, the cost might be too high with real-world services such as those found in a pervasive computing environment. Dummy users might have to control physical objects—opening and closing doors, for example—or purchase services with electronic cash. Furthermore, realistic dummy user movements are much more difficult to construct than dummy messages.

Figure 4: Directions for further research, dummy users [20]

The quote also explicitly recognizes the challenges of creating realistic dummies. One paper [11] also reported this problem when trying to use a trajectory database of vehicle movements. No literature to date, except [50], has examined the possibility of using "live" geographical coordinates for LBS FPL queries, yet these coordinates are readily available to vehicular LBS users.

At least six previous works have studied the use of dummy events in protecting location privacy. Some [11] suggest dummy vehicle locations be generated by adding noise to traces from a trip planner. Another [12] proposes generating dummy locations in the same neighborhood as the genuine current location. Still another [13] recommends generating dummy locations entire trajectories, full trips, using algorithms which offer realistic vehicular mobility modeling and derive positions from databases of previously-recorded real driver locations. One paper [14] advocates dummy location generation using either a local grid called a virtual grid, similar to [12], or a virtual circle, which ensures *k*,*s*-privacy. (A location is said to possess *k*,*s*-privacy when it contains k users in an area smaller than s.) The paper in [15] recommends a scheme which randomly generates dummy locations rotating with movement patterns that consistent with observed human movement, though not vehicle movement. Researchers in [16] study dummies in packet routing, not directly relevant to this study.

Table 2: Location privacy by dummy event, literature overview

Authors	Methods	Category
You, Peng and Lee (2007) [15]	Random trajectory	Spatial shift
Lu, Jensen and Yiu (2008) [14]	Virtual grid, virtual circle	Spatial shift
Chow and Golle (2009) [11]	Google Maps poly line	Trajectory database
Kido, Yanagisawa and Satoh (2009) [12]	Moving in a neighborhood	Spatial shift
Krumm (2009) [13]	Data gathered from GPS receivers, modified with noise	Trajectory database
Alnahash, et. al. (ASEE, 2014) [51]	Random trajectory confined to road grid	Spatial shift
Corser, et. al. (IEEE, 2014) [50]	"Live" dummies generated by active vehicles	Active decoy

None of these studies except [13] generate dummies from realistic vehicular mobility models. None consider the threat model in which LBS administrators collude with RSU administrators. None use active decoys, i.e. false locations (dummy events) of real vehicle locations in real time transmitted by vehicles other than the target vehicle. Besides the EPZ model [28] and PBD model [50] we are aware of no study to date which has examined the deanonymization of endpoints in VANETs under LBS/RSU collusion.

4. Conclusion

In sum, terminology has been used inconsistently in vehicle network privacy literature and this paper presented a standard taxonomy. Further, based on a review of the literature it appears a new body of research is needed to fill the present gap between desired vehicle network properties and existing solutions. It is possible that some location privacy methods, such as dummy events, may be more applicable than originally assumed and should be investigated further.

5. References

[1] Bush, I. (2013, Feb 25). GM, AT&T readying in-vehicle wi-fi. http://philadelphia.cbslocal.com/2013/02/25/gm-att-readying-in-vehicle-wi-fi/

[2] Johnson, L. (2012, Oct 31). Location-based services to bring in $4b revenue in 2012: study. http://www.mobilemarketer.com/cms/news/research/14115.htmlhttp://www.mobilemarketer.com/cms/news/research/14115.html

[3] Koslowski, T. (2012, Jan 3). Your connected vehicle is arriving. http://www.technologyreview.com/news/426523/your-connected-vehicle-is-arriving/

[4] IEEE Standard for Wireless Access in Vehicular Environments Security Services for Applications and Management Messages," IEEE Std 1609.2-2013 (Revision of IEEE Std 1609.2-2006) , vol., no., pp.1,289, April 26 2013, doi: 10.1109/IEEESTD.2013.6509896

[5] Kenney, John B. "Dedicated short-range communications (DSRC) standards in the United States." Proceedings of the IEEE 99.7 (2011): 1162-1182.

[6] Corser, G., Fu, H., Shu, T., D'Errico, P., Ma, W. (2013). "Endpoint Protection Zone (EPZ): Protecting LBS User Location Privacy Against Deanonymization and Collusion in Vehicular Networks." The 2nd International Conference on Connected Vehicles & Expo (ICCVE 2013).

[7] Sampigethaya, K., Huang, L., Li, M., Poovendran, R., Matsuura, K., & Sezaki, K. (2005). CARAVAN: Providing location privacy for VANET. Washington Univ Seattle Dept Of Electrical Engineering.

[8] Guo, J., Baugh, J. P., & Wang, S. (2007, May). A group signature based secure and privacy-preserving vehicular communication framework. In 2007 Mobile Networking for Vehicular Environments (pp. 103-108). IEEE.

[9] Sun, Y., Lu, R., Lin, X., Shen, X., & Su, J. (2010). An efficient pseudonymous authentication scheme with strong privacy preservation for vehicular communications. Vehicular Technology, IEEE Transactions on, 59(7), 3589-3603.

[10] Shokri, R., Freudiger, J., & Hubaux, J. P. (2010). A unified framework for location privacy. 3rd Hot Topics in Privacy Enhancing Technologies (HotPETs).

[11] Chow, R., & Golle, P. (2009, November). Faking contextual data for fun, profit, and privacy. In Proceedings of the 8th ACM workshop on Privacy in the electronic society (pp. 105-108). ACM.

[12] Kido, H., Yanagisawa, Y., & Satoh, T. (2005, July). An anonymous communication technique using dummies for location-based services. In Pervasive Services, 2005. ICPS'05. Proceedings. International Conference on (pp. 88-97). IEEE.

[13] Krumm, J. (2009). Realistic driving trips for location privacy. In Pervasive Computing (pp. 25-41). Springer Berlin Heidelberg.

[14] Lu, H., Jensen, C. S., & Yiu, M. L. (2008, June). Pad: Privacy-area aware, dummy-based location privacy in mobile services. In Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access (pp. 16-23). ACM.

[15] You, T. H., Peng, W. C., & Lee, W. C. (2007, May). Protecting moving trajectories with dummies. In Mobile Data Management, 2007 International Conference on (pp. 278-282). IEEE.

[16] Yang, Q., Lim, A., Ruan, X., & Qin, X. (2010, December). Location privacy protection in contention based forwarding for VANETs. In Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE (pp. 1-5). IEEE.

[17] Machanavajjhala, A., Kifer, D., Gehrke, J., & Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. ACM Transactions on Knowledge Discovery from Data (TKDD), 1(1), 3.

[18] Li, N., Li, T., & Venkatasubramanian, S. (2007, April). t-closeness: Privacy beyond k-anonymity and l-diversity. In Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on (pp. 106-115). IEEE.

[19] Dwork, C. (2006). Differential privacy. In Automata, languages and programming (pp. 1-12). Springer Berlin Heidelberg.

[20] Beresford, A. R., & Stajano, F. (2003). Location privacy in pervasive computing. Pervasive Computing, IEEE, 2(1), 46-55.

[21] Uzcategui, R.; Acosta-Marum, G., "Wave: A tutorial," Communications Magazine, IEEE, vol.47, no.5, pp.126,133, May 2009, doi: 10.1109/MCOM.2009.4939288

[22] Harri, J., Filali, F., & Bonnet, C. (2009). Mobility models for vehicular ad hoc networks: a survey and taxonomy. Communications Surveys & Tutorials, IEEE, 11(4), 19-41.

[23] Baumann, R., Legendre, F., & Sommer, P. (2008, May). Generic mobility simulation framework (GMSF). In Proceedings of the 1st ACM SIGMOBILE workshop on Mobility models (pp. 49-56). ACM.

[24] http://gmsf.sourceforge.net/

[25] http://www.lst.inf.ethz.ch/research/ad-hoc/car-traces

[26] Sweeney, Latanya. "k-anonymity: A model for protecting privacy." International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 10.05 (2002): 557-570.

[27] Gruteser, M., & Grunwald, D. (2003). Anonymous usage of location-based services through spatial and temporal cloaking. In Proceedings of the 1st international conference on Mobile systems, applications and services (pp. 31-42). ACM.

[28] Corser, G., Fu, H., Shu, T. D'Errico, P., Ma, W. (2013, December). Endpoint Protection Zone (EPZ): Protecting LBS user location privacy against deanonymization and collusion in vehicular networks. In Second International Conference on Connected Vehicles & Expo. IEEE.

[29] Glassbrenner, D. (2003). Estimating the lives saved by safety belts and air bags. US DOT, National Highway Traffic Safety Administration , page, 5, 12.

[30] http://icsw.nhtsa.gov/safercar/ConnectedVehicles/pages/v2v.html.

[31] PBS, Feds to announce decision on new automobile safety technology. Feb 3, 2014. http://www.pbs.org/newshour/rundown/feds-announce-decision-new-automobile-safety-technology/

[32] Dedicated Short Range Communications (DSRC) Message Set Dictionary, SAE Std J2735, SAE International,DSRC committee, 2009. SAE

[33] Raya, M., & Hubaux, J. P. (2005, September). The security of VANETs. In Proceedings of the 2nd ACM international workshop on Vehicular ad hoc networks (pp. 93-94). ACM.

[34] TS3290/00A. On-Board Unit. Kapsch. Accessed 2014-04-23. https://www.kapsch.net/ktc/downloads/datasheets/in-vehicle/5-8/Kapsch-KTC-DS-OBU-TS3290_00A?lang=en-US

[35] Lee, E. K., Yang, S., Oh, S. Y., & Gerla, M. (2009, October). RF-GPS: RFID assisted localization in VANETs. In Mobile Adhoc and Sensor Systems, 2009. MASS'09. IEEE 6th International Conference on (pp. 621-626). IEEE.

[36] Bohlooli, A., & Jamshidi, K. (2012). A GPS-free method for vehicle future movement directions prediction using SOM for VANET. Applied Intelligence, 36(3), 685-697.

[37] http://www-fars.nhtsa.dot.gov/Main/index.aspx

[38] http://www.fhwa.dot.gov/policyinformation/travel_monitoring/tvt.cfm

[39] Haas, J. J., Hu, Y. C., & Laberteaux, K. P. (2011). Efficient certificate revocation list organization and distribution. Selected Areas in Communications, IEEE Journal on, 29(3), 595-604.

[40] http://www.merriam-webster.com/dictionary/anonymous

[41] http://www.merriam-webster.com/dictionary/pseudonymous

[42] Dıaz, C., Claessens, J., Seys, S., & Preneel, B. (2002, May). Information theory and anonymity. In Proceedings of the 23rd Symposium on Information Theory in the Benelux (pp. 179-186).

[43] Shokri, R., Troncoso, C., Diaz, C., Freudiger, J., & Hubaux, J. P. (2010, October). Unraveling an old cloak: k-anonymity for location privacy. In Proceedings of the 9th annual ACM workshop on Privacy in the electronic society (pp. 115-118). ACM.

[44] Coull, S. E., Wright, C. V., Keromytis, A. D., Monrose, F., & Reiter, M. K. (2008). Taming the devil: Techniques for evaluating anonymized network data. In Network and Distributed System Security Symposium 2008: February 10-13, 2008, San Diego, California: Proceedings (pp. 125-135). Internet Society.

[45] Xiao, X., & Tao, Y. (2007, June). M-invariance: towards privacy preserving re-publication of dynamic datasets. In Proceedings of the 2007 ACM SIGMOD international conference on Management of data (pp. 689-700). ACM.

[46] Dewri, R., Ray, I., & Whitley, D. (2010, May). Query m-invariance: Preventing query disclosures in continuous location-based services. In Mobile Data Management (MDM), 2010 Eleventh International Conference on (pp. 95-104). IEEE.

[47] Kelly, D. J., Raines, R. A., Grimaila, M. R., Baldwin, R. O., & Mullins, B. E. (2008, October). A survey of state-of-the-art in anonymity metrics. In Proceedings of the 1st ACM workshop on Network data anonymization (pp. 31-40). ACM.

[48] Park, H., & Kent, S. (2009). Traceable anonymous certificate.

[49] Chaum, D. (1983, January). Blind signatures for untraceable payments. In Advances in cryptology (pp. 199-203). Springer US.

[50] Corser, G., Fu, H., Shu, T. D'Errico, P., Ma, W., Leng, S., Zhu, Y. (2014, June). Privacy-by-Decoy: Protecting Location Privacy Against Collusion and Deanonymization in Vehicular Location Based Services. In 2014 IEEE Intelligent Vehicles Symposium. IEEE. Dearborn, MI.

[51] Alnahash, N., Corser, G., Fu, H. (2014, April). Protecting Vehicle Privacy using Dummy Events. In 2014 American Society For Engineering Education North Central Section Conference (ASEE NCS 2014).