Privacy Endangerment from Protocol Data Sets in VANETs and Countermeasures

Sebastian Bittl, Arturo A. Gonzalez

Fraunhofer ESK, 80686 Munich, Germany
{sebastian.bittl, arturo.gonzalez}@esk.fraunhofer.de

Abstract. Wireless vehicular networks are about to be deployed within the next years. Important progress towards practical usage of such networks is being made by standardization in Europe and the USA. Thereby, one of the core concerns is privacy of vehicles and their drivers, especially in Europe. Prior work has regarded only a small sub-set of the information exposed by current standards to an attacker for vehicle tracking. Thus, we take a close look on the data contained on different protocol layers of an ETSI ITS system. We find that much data is very distinctive and can be used to identify static vehicle properties such as manufacturer or even model. We call these data sets volatile constant data. Its presence is shown to greatly reduce usability of formerly proposed cooperative pseudonym switching strategies. Thereby, a privacy metric called vehicular uniqueness is introduced. The provided analysis shows that more constraints have to be applied for selecting appropriate cooperation partners for pseudonym switching, which significantly reduces their availability. Therefore, current techniques cannot provide the level of privacy defined in VANET standards. Suggestions for improving the data sets used by security entity and facility layer of ETSI ITS are given to limit the impact of the found issues. Effectiveness of the proposed mechanisms is shown in the provided evaluation.

Keywords: VANET. ETSI ITS. Privacy. Security.

1 Introduction

Wireless intelligent transport systems (ITS) are about to enter the mass market in upcoming years. Important examples are ETSI ITS in Europe [2] and WAVE in the USA [20]. Thus, these systems' security and privacy aspects are gaining increased attention. Thereby, the possibility to track vehicles is a core point of concern, especially in Europe [26]. Many approaches for realizing such tracking exist. Typically, such attacks use the temporarily fixed pseudonym certificate, used by vehicles to authenticate their broadcast messages. However, higher level protocol information, e.g., identifiers or current position and velocity of a vehicle, is regarded for this purpose as well [17].

Many studies have shown the possibility to track vehicles in ITS systems based on the mentioned data sets, e.g. [28]. Tracked vehicle movement paths

can be well correlated to homes of their drivers [18]. Thus, tracking of vehicles limits privacy of drivers. Therefore, a number of countermeasures has been published. These include context aware pseudonym changes [17] and time synchronized pseudonym switching [29]. Unfortunately, these mechanisms require the exchange of further messages between vehicles cooperating during the pseudonym change. This clearly increases the already significant overhead introduced by security mechanisms (see e.g., [13]). Moreover, none of these works studies the influence of metadata contained in the security envelope of current ETSI ITS and WAVE systems on the privacy of vehicles. Additionally, only a fraction of the vast number of data fields from higher level applications is taken into consideration in this prior work.

Thus, a first look on the privacy impact of such additional data fields has been provided in [12]. This work extends the one given in [12], especially regarding recently updated ETSI ITS standards and a significantly increased evaluation of suggested privacy improvement mechanisms.

Our contribution focuses on the influence on the privacy of information broadcast by vehicles in ETSI ITS conforming VANETs. Thereby, we especially study the metadata contained in the security envelope of broadcast messages apart from the used pseudonym. Furthermore, we take a close look on the high number of data sets used by higher level protocols regarding their possibility to ruin the privacy efforts taken elsewhere. For the sake of compactness we focus the study of the ETSI ITS facility layer on Cooperative Awareness Messages (CAMs). However, much similarity between ETSI ITS and WAVE exists on the different protocol layers. Thus, we especially point out the cases which also apply for WAVE based systems.

The further outline is as follows. Section 2 reviews related work. Afterwards, Section 3 provides the in detail study of the impact of individual data fields on privacy of broadcasting vehicles. In Section 4, the achieved results are used to determine a metric for vehicle uniqueness within its vehicular environment. Finally, a conclusion is provided in Section 6 alongside with possible topics of future work.

2 Related work and Attacker Model

Recent work on privacy in vehicular area networks (VANETs) or intelligent transport systems (ITSs) includes [11, 15–17, 25, 28, 29]. Basically, privacy in such networks relies on a pseudonym scheme which changes the identifiers (IDs) of a vehicle (or ITS-station (ITS-S)) on all protocol layers on a regular basis to avoid tracking [26]. A general overview of security and privacy mechanisms in VANETs is given in [24, 26, 27].

Both ETSI ITS and WAVE do not use dedicated messages to distribute security related data sets like pseudonym certificates. Instead piggybacking of such information on cyclically (CAM, BSM) or on demand distributed (DENM) messages is used [3, 9].

There are mainly two kinds of attacks on privacy in VANETs. Simple attacks just use identifiers like the station ID and a very limited set of additional information about the ITS-S, typically only the vehicle position. Advanced attacks include more context information for tracking, e.g., behavior of other vehicles [17, 28]. Bayesian traffic analysis to re-identify vehicles after cooperative pseudonym changes is proposed in [14]. Thereby, it has been shown that simple pseudonym change, like in ETSI ITS and WAVE, cannot avoid tracking. The probability of two (or even more) vehicles changing their pseudonym in close vicinity just by chance, confusing an attacker, is just to small.

Many approaches to confuse an attacker trying to track vehicles have been proposed. A simple pseudonym scheme based on per-trip certificate usage is given in [11]. Advanced pseudonym switching schemes apply concepts like *Mix-Zones* [25], silent periods, *SLOW* [15], context aware pseudonym changes [17] and time synchronized pseudonym switching [29] (see also [28] and references within). A common requirement of all these concepts is that vehicles must find indistinguishable partners in their vicinity with whom they cooperate to perform a secure pseudonym change. All vehicles should change all of their identification parameters together to confuse the attacker [26]. However, we show that finding such partners is quite unlikely to happen in VANETs using current ETSI ITS and WAVE standards.

A commonly assumed attacker model is the global passive adversary [28]. This passive attacker can monitor all messages in the whole ITS system. This model is also assumed in the following.

Even advanced attacks from prior work (e.g., [28, 29]) have so far not included usage of the biggest share of metadata from the security envelope and higher protocol level data from cyclic messages in VANETs following current standards. Thus, we study the usability of these data for more advanced attacks.

Properties of the studied standards from the ETSI ITS and WAVE frameworks are explained in the next Section 3, alongside with their impact on privacy aspects.

3 Data for Vehicle Identification

ETSI ITS and WAVE use Cooperative Awareness Messages (CAMs) and Basic Safety Messages (BSMs) for the main data exchange in their VANET systems, respectively. Therefore, our focus is on the contents of theses messages on the different protocol layers.

When looking for possible privacy issues regarding vehicle tracking, the core focus is on data within messages which differs for different groups of vehicles, but is also constant for the individual vehicle for a long time. One example is vehicle dimensions which are identical for all vehicles of the same model but different for other models with high probability. We call that kind of data *volatile constant data*. Thereby, volatile is meant in the sense of the data being accessible by all receivers and constant with respect to a long time period.

Within typical traffic scenarios many different vehicle types and models are present in the vicinity of a vehicle intending to perform a secure pseudonym change. To do so, the above described cooperative pseudonym strategies select partners whose broadcast information is as similar as possible to confuse the attacker. The presence of volatile constant data clearly makes it less probable to find such proper partners leading to possibly insecure pseudonym changes. Thus, the presence of such data should be avoided as far as possible.

Current standards bind the lifetime of MAC address, network layer address and station ID of the facility layer to the one of the pseudonym. This means, once the pseudonym gets changed the other identifiers get changed, too. Therefore, an attacker cannot profit from looking on more than one of these identifiers at once as they all provide the same temporarily valid information. Moreover, for the simple case of single hop broadcast, like it is used for CAMs and BSMs, the network and access layer only add information to the transmitted messages which cannot be used to track their senders.

In the next section metadata from the security envelope will be studied. The impact of data from access and network layer is looked at in Section 3.2. Afterwards, the content of CAMs at the facility layer will be discussed.

3.1 Metadata in Security Envelope

The security envelope is used to secure content from the facility and network layer protocols by embedding them into a dedicated header and trailer, each consisting of different sub-parts. Thereby, content handed over to the security entity is treated in accordance to a so called security profile. These profiles determine the required header fields as well as the used cryptographic techniques, which can be digitally signing and/or encryption. The definition of the security envelope is quite similar in ETSI ITS [9] and WAVE [3].

In ETSI ITS the sets of mandatory header fields for security profiles *CAM* and *Generic* are subsets of the one for *DENM* (used for Decentralized Environment Notification Messages (DENMs)). Thereby, the location stamp in profiles *DENM* and *Generic* carries the same information as the vehicle position inside a CAM [5, 7]. Thus, this field is not discussed separately and the reader is referred to Section 3.3 for details.

We focus the further discussion on mandatory header fields from the CAM security profile. Privacy issues resulting from such fields are more severe than those from optional ones, as these can be simply skipped in practical implementations. In contrast, to fix issues regarding mandatory fields the standard has to be changed. Furthermore, [23] suggests to remove the possible inclusion of optional fields. We support this proposal, as differing combinations of data sets in the envelope by different implementations clearly give an attacker a possibility to easily distinguish vehicles independently from their pseudonyms.

The following sections discuss the different header fields' privacy implications in detail.

Protocol Version The used protocol version will be constant for all vehicles at the beginning of the deployment phase. However, over time it is very likely that multiple versions will be present in VANETs. As this value is constant for an individual vehicle over a long time, it is clearly volatile constant data. Thus, the presence of many different versions should be avoided even if they are otherwise compatible.

Signer Info The signer information of a message may hold different contents. Thus the available information for the attacker differs. However, one can always uniquely determine the signer (and sender) of the message. Therefore, this is often called the pseudonym ID of the sender. In both CAMs and BSMs the field's content can either be the hash of the used pseudonym certificate (PSC) or the full certificate [9, 3]. Both systems use cyclic inclusion of the full certificate every 0.5 or 1 second, respectively. In case of security profiles *DENM* and *Generic* the full certificate is always present [9].

In case of an included PSC within the security envelope the following data is available to the receiver [3, 9].

Signer Info of Pseudonym Certificates A signer info field in a PSC identifies its signer, which is an authorization authority (AA). This can be done either by a hash digest or by the full AA certificate (AAC). Both uniquely identify the AA. Current standards allow for a possible multitude of such entities to exist. In practice this will be probably done by the car manufacturers (OEMs). However, this leads to a privacy issue as the signer information is volatile constant data. An attacker can directly determine the OEM and use this to distinguish PSCs and thereby vehicles. PSCs signed by different AAs are very unlikely to be used by the same vehicle and a vehicle will very likely use only PSCs issued by the same AA. Clearly, vehicles of low volume OEMs will be particularly vulnerable.

To limit the usability of the AA's identity for an attacker one can think of mainly two countermeasures. Firstly, one could increase the number of AA certificates and make a single AA use a multitude of them. Thereby, the effort for an attacker to keep track of all certificates would increase. However, this would significantly increase the effort for AA certificate distribution to all ITS-S for a small security gain.

Secondly, one could limit the number of AAs. An ideal choice would be to have only one AA. This would clearly resolve the above described privacy issue completely, as an attacker cannot distinguish vehicles based on their used AA anymore. To implement this, OEMs would have to cooperate and use a common AA. As they plan to establish a common root certificate authority (CA) for Europe, this seems to be a usable approach. In order to limit the number of PSCs signed by a single AA certificate, one could significantly limit its lifetime. New ones can be deployed together with PSC updates.

Additionally, one should coordinate the lifetime of an AA's certificate with the lifetime of its issued PSCs. Thereby, any possibility to distinguish PSCs based on their signing AA should be ruled out. Moreover, the number of AA certificates to be stored securely inside the vehicles is kept (very) low.

The privacy gain from using a single AA is evaluated in Section 5.2 in more detail.

Validity Restriction The mandatory validity restriction of PSCs is a limited validity period. It is determined by a start and end time stamp. Both are used with an accuracy of one second. The PSC distribution scheme described in [3] and [1] defines that PSCs are delivered from an AA to an ITS-S upon request of the ITS-S. The remaining details are implementation specific, as they are not covered by the standard. However, a possible pitfall for privacy of pseudonym users exists which is caused by the mentioned time stamps.

This pseudonym usage privacy issue arises from the planned way of (re-)using PSCs in Europe. Thereby, each vehicle uses a pool of PSCs which are (re-)used until the full pool gets updated [11, 28]. The update period will probably be in the order of months.

A different approach is described in [20] for WAVE in the USA. Thereby, each PSC is only used once and the validity period is the order of minutes. However, this approach is not favored by manufacturers and maintainers of ITS-Ss, as it introduces significant overhead in the ITS system for PCS distribution. Either vehicles require frequent, reliable connections to the AA (or pseudonym certificate authority (PCA) [20]) or a huge buffer filled with PSCs for future use. Even if the initially proposed validity period of five minutes gets doubled, this would still require a maximum amount of 144 PSCs per day. To protect the buffered PSCs, these have to be stored in secure memory, e.g., inside a Hardware Security Module (HSM). However, adding more memory to an HSM significantly increases its price. Moreover, many issued PSCs will stay unused as their validity period elapses while the vehicle is not in use. One would have to know the usage times of each vehicle in advance to avoid that, which is hardly practicable. Thus, the approach from [20], while providing good privacy, probably bears too much overhead for large scale deployment.

An alternative approach for securing re-usage of PSCs is discussed in the following.

There are mainly two approaches for PSC generation inside the AA. Either the AA generates the PSCs upon request or the AA keeps track of the expiration of its users' PSCs to generate new ones in advance. In both cases a straight forward implementation would take the same time stamp (e.g., the current time at the AA) and use it as the common start validity time stamp of the signed PSCs. However, this means that all PSCs of a set delivered to an ITS-S have a very similar (or even the same) start validity time stamp. Thereby, making this information volatile constant data. Furthermore, this time stamp will be different with a very high probability for most cars as there is no timed synchronization of PSC requests.

The PSC users have no possibility to protect themselves against an attacker using validity time stamps for tracking them, as they cannot change the content of a PSC without invalidating its signature. Therefore, countermeasures have to be taken within AAs.

A straight forward solution would be to discretized the time stamps defining the validity period of PSCs. For example, all PSCs issued in one month could receive the start of this month as their start validity time stamp. The longer the discretization steps, the more vehicles will receive a set of PSCs with the same validity period. Consequently, the probability that multiple vehicles with common values in these data fields meet on the street increases removing the possibility to distinguish them.

Subject Attribute The subject attribute field holds the subject type and public key of the PSC. This key is randomly generated and the subject type is fixed for all PSCs. Thus, there is no possibility to link PSCs based on this data set.

Subject Info The subject info field holds a fixed value for all PSCs. Thus, it provides no possibility to track vehicles.

Generation Time The generation time is individual for each message. However, the time difference between two sequential messages is clearly defined by the standard. Neither ETSI ITS nor WAVE define any change to the sending interval before or after a pseudonym change.

A common assumption is that clocks of ITS-S are well synchronized using GPS [29]. Thus, time intervals between message generation of individual cars should be quite stable. Additionally, inside a group of cars the generation times of messages should be randomly distributed leading to an even distribution of used time stamps. These time stamps are generated and transmitted with microsecond resolution [9]. Hence, collisions in this data field which could confuse an attacker are unlikely. Thus, an attacker can track vehicles just based on the generation time of their messages with high probability.

In case of BSMs the sending interval is fixed. For CAMs, it is determined by multiple parameters and can be in the range from 1 to 10 Hz. However, the current interval can be found in the transmitted CAM itself [7]. This allows the attacker to easily use this information to avoid being confused by the variable sending interval of CAMs.

Furthermore, the time step is set at the network layer. Hence, tracking capabilities of the attacker are not limited by the lower layer channel access mechanism. For example, the actual sending time being somehow randomized by the probabilistic lower layer CSMA-CA scheme, used within ITS-G5 and IEEE 802.11p, does not confuse the attacker.

We propose two solutions to overcome the described vulnerability. Both require the cooperating vehicles to use the same sending frequency before and after the pseudonym change for a minimum time span, e.g., one second. Firstly, one could reduce the accuracy of the generation time to the maximum transmission interval being 100 ms for BSMs and 1s for CAMs. The security entity does not need to determine the sequence of received messages according to standards. Moreover, the validity time spans of PSCs are also given with full second resolution. Therefore, currently there is no need to use a high precision time stamp for

the generation time of type *Time64* and it should be substituted by the lower resolution *Time32* type. A side effect would be a reduction in the size of the security envelope by four bytes [9].

For the second solution, immediately after the pseudonym change the next sending must be delayed by a random waiting time. The true random number generator required for ECDSA signatures could be used to obtain it.

The length of the random waiting time should be in the order of the normal time difference between two successive transmissions. For example, for BSMs it would be between zero and 100 ms. Consequently, the attacker cannot determine the next generation time and gets confused. The impact on higher level layers, e.g., applications, should be low. From their perspective a maximum delay looks just like one missed message from the other vehicle.

ITS Application Identifier The content of the ITS Application Identifier (AID) field identifies target application as well as message type (e.g., CAM) and thereby also the security profile used for the given message. In case all vehicles monitored by an attacker only send the same type of message, e.g., CAMs, he cannot discriminate the sender based on this data. However, future extensions of VANET communication may lead to a multitude of different messages sent by ITS-Ss. In case these send differing sets of messages, e.g., due to differing available applications, an attacker can distinguish them and this information is volatile constant data.

A receiving ITS-S should use the ITS AID value to check whether the sender's PSC allows to sign this particular type of message. However, the security gain of this mechanism is very low. The reason for this is that the security entity cannot check if the real payload really corresponds to the received ITS AID value. Thus, an attacker who can assemble an arbitrary message can put unauthorized content, e.g., a DENM inside the network layer payload. Then, he can have it signed with the CAM security profile by the security entity with a PSC being only valid for CAMs. Thereby, the receiver's security entity will accept the incoming message as being properly signed.

Instead, the attack has to be detected at the application or facility layer. The receiving entities at these layers have to determine the actual ITS AID of the received message and need to check whether the sender's certificate holds the required privileges.

Furthermore, the attacker is not required to have direct control over the PSC to carry out the above described attack. It is sufficient to have control over the interface to the security entity, as it has no possibility to check the content of the payload it wraps into the security envelope. Thus, we consider the security gain of the presence of the ITS AID field as neglectable.

Moreover, the security entity does not need to distinguish different message types sharing the same security profile. Therefore, the ITS AID field should be removed from the security envelope, as it only adds overhead to it. Instead, one should limit the contained information to the used security profile, as done in the preceeding standard version [5].

Certificate Request List An ITS-S requests up to six unknown certificates (PSCs or AACs) by using the least three bytes of their hash values. Standards are unclear about when to remove entries from the request list [9, 3]. It should be flushed after a pseudonym change, as the current state of certificates required by an ITS-S can be expected to be highly discriminative between ITS-Ss.

Trailer Field There is only one type of trailer field in the standards. It holds metadata for interpreting the digital signature as well as the signature itself. Most parts of the trailer are fixed and the signature of multiple messages can only be linked together with the help of the respective public key. Therefore, the signature does not carry any additional privacy related information compared to the public key in the corresponding PSC (see Section 3.1 above).

Moreover, the encoding of the used ECC (elliptic curve cryptography) point may vary in general, but is probably constant for a particular vehicle. There are four options for the ECC point type field in the standard, with the core difference being enabled or disabled ECC point compression. With both choices used, this information is volatile constant data. In the worst case, with only two cars in a group and both using a different ECC point type, this information is already enough to render any pseudonym change useless. Thus, the standard should only allow only one option to be used. For other reasons to do so see [23].

3.2 Data from Access and Network Layer

The identifiers of the access layer (i.e., MAC address) and the network layer (so called GeoNetworking address) are coupled to the hash value of the currently used PSC. Thus, their impact on privacy is the same as outlined in Section 3.1 for the signer identifier field inside the security envelope.

3.3 Data from Facility Layer

The CAM is defined as a deeply nested data structure holding mandatory and optional data sets [8]. Thereby, an *ItsPduHeader* and a *CoopAwareness* field are present on the top level. The simple *ItsPduHeader* only holds basic information like the protocol version, message id and station id. These fields hold the same information as their respective counterparts in the security envelope. Therefore, their impact on privacy aspects is the same as for those data sets already described in Section 3.1.

The *CoopAwareness* field has two parts being the current generation interval (usable by an attacker as described in Section 3.1) and the *CamParameters* field consisting of several different so called containers (i.e. dedicated data sets). These are described in detail in the following.

Basic Container The always present basic container holds the components station type and reference position.

Station Type The station type associates the vehicle to some generic class, e.g., passenger car or light truck. This unchanging information is clearly volatile constant data.

Reference Position The current position of the ITS-S measured at the vehicle's reference point (see [4]) is available in each CAM. Prior work already showed that this information can be used to bypass simple pseudonym changes [17, 29]. Therefore, the advanced pseudonym switching strategies suggested in these references should be used.

High Frequency Container The high frequency container is part of every CAM. In case of an ITS-S being a vehicle the only used sub-part is a basic vehicle container. Parameters of the vehicle's current movement are given in this data set. These include heading, speed and driving direction. All these values can be used for advanced vehicle tracking [17, 29]. However, the remaining data inside this container has not been regarded in prior work.

Dimensions The vehicle's dimensions length and width are given. According to [7] the resolution is set to 0.1 meters. This value stays constant during one journey of a vehicle and thus it has to be regarded as volatile constant data. It is possible that the length of a vehicle changes from one journey to another, e.g., by extending it with a trailer. However, this is rare in practice especially for passenger cars.

To evaluate privacy aspects of broadcasting a vehicle's dimensions, we determined the number of different currently sold vehicle models in Germany. Then, we assigned them to the individual discretization steps of vehicle length and width. We took publicly available data from the German Kraftfahrt-Bundesamt [21] to obtain the share of different vehicle types, separated into OEMs and their models, on the overall traffic in Germany caused by new cars. Foreign cars traveling on German roads are excluded from this data set. However, it should still give a reasonable estimate about the distribution of models' dimensions. Moreover, we used public information from the 45 different OEMs present in [21] to obtain the individual dimensions of models.

We find that 73% of all vehicle models share a common combination of width and length with at least one other model. These cars have a market share of 75%. Thus, for a share of 25% one can determine the model directly given its discretized dimensions. Even the most populated set of vehicles with length 4.3 m and width 2.0 m includes only 17% of all cars.

Thus, distribution of vehicle dimensions clearly decreases the probability to find proper (i.e., indistinguishable) partners for a cooperative pseudonym change. Further discretization of the values to, e.g., 0.3 m would significantly improve the situation for many vehicles but can still not help the ones with outstanding dimensions and/or low penetration rates.

Dynamics The parameters longitudinal acceleration, curvature (consists of curvature value and confidence), curvature calculation mode and yaw rate are in-

cluded in the high frequency container. Thereby, the curvature calculation mode is again a value which is unlikely to change for an individual vehicle and may differ for different vehicles. Therefore, it should be regarded as volatile constant data.

The remaining values model a vehicle's trajectory. Many approaches for modeling and predicting such trajectories exists, e.g., [10, 19]. In case of pure tracking, i.e., no realtime interaction between attacker and vehicles, the attacker does not need to process the information in realtime. Thus, he can use computationally expensive but accurate and complex movement models. As we have seen above, the attacker can determine either the vehicle type directly or a group of possible vehicle types. This information can be used to tune the parameters of a movement model making it very accurate. Moreover, the prediction must only work well for a short time span as the CAM generation rate is at most one second.

To evaluate the impact of using an advanced movement model on the attackers ability to track vehicles one should use data obtained from real test drives instead of pure simulator output. This is because simulators like the well known SUMO use a predefined vehicle model. Therefore, tracking these simulated vehicles with a model which fits the one used to generate their movement will probably yield unrealistically high success rates. Further analysis of this issue is beyond the scope of this work and is a subject to future work.

Optional Data Six more data sets may be optionally present in the container. Three of them (steering wheel angle, lateral, vertical acceleration) can be used to improve the movement model described above.

The remaining three values (acceleration control, lane position, performance class) each describe a vehicle's feature. These can be expected to change quite slowly, i.e., they should be regarded as volatile constant data. As all these fields are optional and can be added or removed individually, also the combination of sent data sets may differ between vehicles. Thus, usage of each extra value will increase the change that a particular vehicle uses a unique set of data inside its current vicinity. Thereby, it will strip itself from finding proper partners for a secure pseudonym change.

Optional Containers In addition to the basic and high frequency container, the low frequency container is distributed cyclically, but not in every single CAM. It contains the vehicle role, exterior lights and path history fields. See [8] for details about inclusion rules.

In case of an uncommon vehicle role, e.g., rescue vehicle, the corresponding additional container is present in the CAM. The density of such vehicles in ordinary traffic is usually low. Thus, an attacker can easily track them just based on the presence of their dedicated containers in their CAMs.

Typically, the status of exterior lights changes slowly. Thus, this data set is volatile constant data.

The path history field should obviously be erased when a pseudonym change occurs or the inclusion rate of the container has to be such low that sequentially

sent values of this field cannot be linked. Otherwise the attacker can simply link the pseudonyms based on this data. However, the current standards do not specify such behavior, but it is recommended in [6].

4 Vehicle Uniqueness

Secure pseudonym switching schemes from prior work are based on the assumption that broadcast data cannot be mapped to an individual vehicle except of the changed identifiers. We have shown in Section 3 that this is clearly not the case due to the presence of volatile constant data. To evaluate the impact of our findings on vehicle privacy we introduce the metric of *vehicle uniqueness (VU)*. It measures how much a particular vehicle differs from its vehicular environment regarding data observable by an attacker.

Prior work showed that tracking of vehicles becomes more difficult alongside with higher traffic density and longer distances traveled during a cooperative pseudonym switching maneuver [28]. However, this only holds in case the attacker has no extra information for re-identification of vehicles after a pseudonym change. VU is a metric for the availability of such extra information. In case a vehicle is unique inside the area of pseudonym switching the attacker can always track it, independently of the used pseudonym switching algorithm.

To calculate VU an exposed feature vector $\boldsymbol{e}_i$ holding all available volatile constant data is assigned to each vehicle. Thereby, $i \in \boldsymbol{I}$ relates to a particular vehicle within a group of vehicles $\boldsymbol{I}$ ($|\boldsymbol{I}| \geq 1$) cooperating during a pseudonym change. VU is defined by

$$VU = \mathbf{Pr}\{|\{x|\boldsymbol{e}_x = \boldsymbol{e}_y; x \neq y; x, y \in \boldsymbol{I}\}| = 0\} \ .$$

This means that $VU \in [0; 1]$ is the probability that there is just one car within $\boldsymbol{I}$ having one particular exposed feature vector. Such vehicles are indistinguishable for an attacker in regard to volatile constant data. Thus, these vehicles are proper candidates for performing a cooperative pseudonym change.

In the following, we take three different pseudonym switching schemes into regard. These are

1. uncoordinated pseudonym switching (ETSI ITS and WAVE) with $|\boldsymbol{I}| = 1$ with high probability,
2. mix zones with $|\boldsymbol{I}|$ depending on traffic flow and size of the mix zone and
3. silent periods with $|\boldsymbol{I}|$ depending on traffic flow and length of silent periods.

Within current standards pseudonym changes are uncoordinated, as every ITS-S decides on its own when to perform the change without including information from other ITS-Ss in its decision process. Results of an evaluation using the proposed VU metric are given in Section 5.

5 Evaluation

In the following we take three different system parametrizations into consideration. At first, a system using a multitude of AAs is studied. This resembles the currently planed way of realizing VANETs based on ETSI ITS and WAVE. Secondly, the approach from Section 3.1 for usage of a single AA is studied to show its significant improvement potential on privacy of ITS-Ss. Afterwards, privacy impact of further discretization of vehicle dimensions as suggested in Section 3.3 is studied. Lastly, a summary about results achieved in the given evaluation is provided in Section 5.4.

5.1 Multiple Authorization Authorities

In this section we include the following data into $\boldsymbol{e}_i$:

- AA of PSCs, we assume one AA per OEM and
- vehicle dimensions (see Section 3.3).

We assume that all cars from the same OEM use the same encoding of ECC points (see Section 3.1). Thus, this data does not influence VU in our case and is not regarded further. The rest of the volatile constant data sets from Section 3 are assumed to be identical for all cars. This leads to a best case assumption for privacy of vehicles, i.e., a worst case assumption for the attacker. Moreover, we assume that the probability of two cars within $\boldsymbol{I}$ sharing a common value of $\boldsymbol{e}_i$ ($|\boldsymbol{e}_i| = 3$) only depends on the share of their particular model within the set of all vehicles.

We use the vehicle distribution from [21] to estimate VU. Moreover, an analysis of vehicle dimensions for the models of different OEMs (see also Section 3.3) shows that the data included in $\boldsymbol{e}_i$ allows to uniquely identify the model of a vehicle, e.g., as VW Golf VII, from a single CAM including the PSC. Thus, one can calculate the probability to encounter a vehicle with a particular $\boldsymbol{e}_i$ from the mentioned vehicle distribution data set.

The number of vehicles encountered during a pseudonym switching maneuver $|\boldsymbol{I}|$ is varied by varying the traffic flow (given in $\frac{vehicles}{kilometer}$) and size of mix zones or length of silent periods, respectively. The traffic density is varied from 16 to 45 $\frac{vehicles}{kilometer}$ per lane following [17] to represent low volume traffic as well as a jammed setup. We use the parameter set from [28] for the size of mix zones (25 m - 400 m), length of silent periods (1250ms -20s) and velocity range (0 - 250 $\frac{km}{h}$). The obtained results are given in Figure 1.

Thereby, the *best* case relates to the most common car model. It is obviously the least unique one within the set of all vehicles. However, only about 7.7% of all vehicles can profit from the good results for this model having a high chance to find indistinguishable partners for a cooperative pseudonym change. Moreover, the *worst* case relates to the least common car.

One can see from Figure 1 that the value of $1 - VU$ increases alongside with $|\boldsymbol{I}|$. However, for an *average* vehicle it is very low for all regarded values of $|\boldsymbol{I}|$.

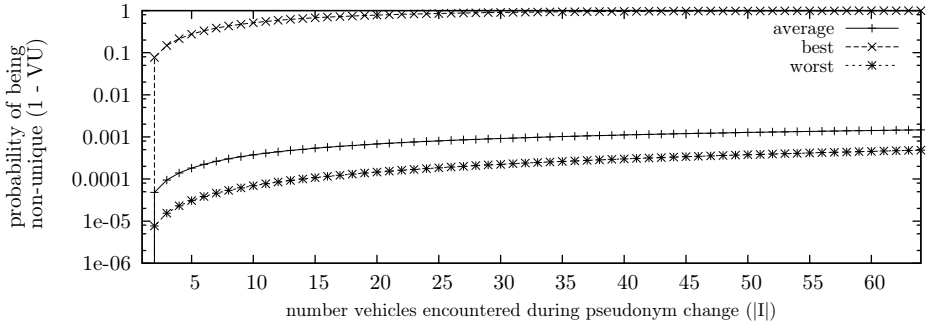

Fig. 1. Vehicle uniqueness during pseudonym change with $|\boldsymbol{e}_i| = 3$.

However, combinations of high velocity and high traffic flow, leading to high values of $|\boldsymbol{I}|$, rarely occur in practice. Thus, VU will exceed 99.9% in most real world scenarios with moderate traffic flow.

This means that the attacker can track those ITS-Ss even after a performed PSC change with more than 99.9% probability. In combination with other techniques from prior work, such as trajectory based tracking, hardly any privacy of vehicles can be expected to remain.

Higher values of $|\boldsymbol{I}|$ than the ones used above would relate to unrealistically high traffic flow or extending the size of mix zones and length of silent periods to values rendering higher level ITS-applications unusable [28]. One should note that, even medium size mix zones have been shown to lead to significant performance degradation of VANET based driver assistance systems [22]. Calculation of VU is independent of the pseudonym switching strategy, but the achievable size of $|\boldsymbol{I}|$ differs. While cooperative PSC switching strategies can adjust it, uncoordinated ones, e.g., from ETSI ITS or WAVE, cannot do so.

The obtained data on vehicle uniqueness shows that the presence of volatile constant data is able to render PSC changes during driving almost useless. An attacker can almost always re-identify vehicles based on this data after the pseudonym change.

5.2 Common Authorization Authority

In order to reduce vehicle uniqueness the usage of a single AA for all ITS-Ss, as proposed in Section 3.1, is considered in the following. Thus, in contrast to Section 5.1 the exposed feature vector $\boldsymbol{e}_i$ only holds vehicle dimensions, i.e., $|\boldsymbol{e}_i| = 2$. The AA identity is no longer present in e_i as it is identical for all ITS-Ss.

The same vehicle distributions and traffic scenarios as in Section 5.1 are used for evaluating the proposed privacy improvement technique. Thereby, well comparability of both systems is ensured.

Obtained results for the system using $|\boldsymbol{e}_i| = 2$ are given in Figure 2.

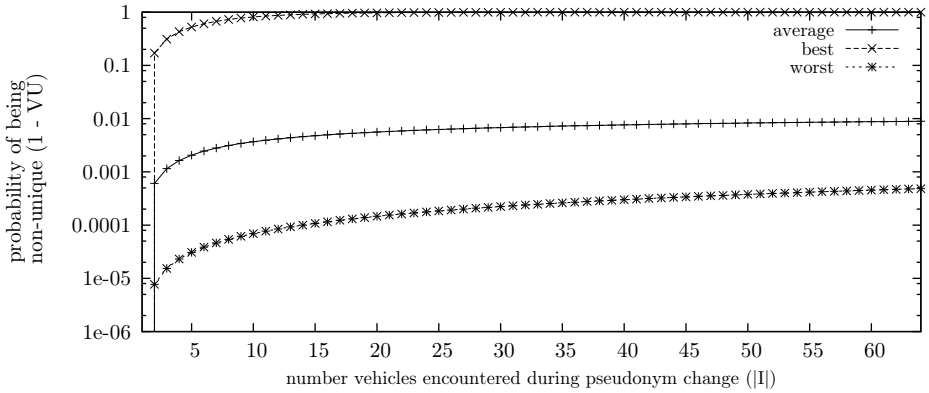

Fig. 2. Vehicle uniqueness during pseudonym change with $|\boldsymbol{e}_i| = 2$ and standardized vehicle dimension's accuracy.

Comparing the results from Figure 2 to the ones from Figure 1, one can see that using a single AA reduces vehicle uniqueness (increasing $1-VU$) by a factor of about 8. Thus, privacy of ITS-Ss is increased. The values for the least common vehicle are unchanged. However, for the most common and average vehicle an improvement of privacy is achieved, although uniqueness of an average vehicle is still quite small. The most common group of indistinguishable vehicle contains about 17.0% of all vehicles.

5.3 Further Discretization of Vehicle Dimensions and Common Authorization Authority

A further mechanism to reduce vehicle uniqueness is to reduce the accuracy of vehicle dimensions included in CAMs, as suggested in Section 3.3. To evaluate its impact the setup described in Section 5.2 is used. Additionally, resolution of vehicle dimensions length and width are further discretized to resolutions of 0.2 m and 0.5 m. With decreasing resolution the data quality available for applications is lowered. However, no detailed requirements regarding this parameter set have been published so far. Thus, future work is required to obtain a tradeoff between privacy and application requirements.

Obtained results for the system using $|\boldsymbol{e}_i| = 2$ together with lowered accuracy of vehicle dimensions are given in Figure 3.

One can see from the comparison of Figures 2 and 3 that lowering the resolution of vehicle dimensions, as given in CAMs, significantly decreases vehicle uniqueness. Thus, privacy of vehicles is well increased. The most common group of vehicles holds about 20.8% and 50.8% of all vehicles for resolutions of 0.2 m and 0.5 m, respectively.

Usage of lowered vehicle dimension's resolution without a common AA would be possible. However, we find that the increase in vehicle privacy is quite low even when using a 0.5 m resolution. The reason for this is that, within the fleet

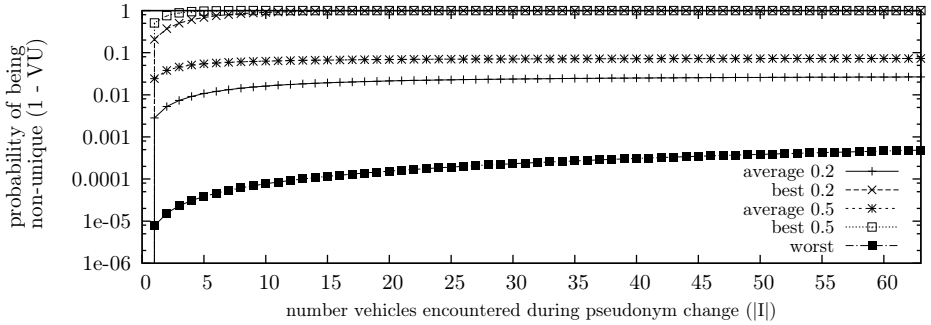

Fig. 3. Vehicle uniqueness during pseudonym change with $|\boldsymbol{e}_i| = 2$ and lowered vehicle dimension's accuracy.

of a single vehicle manufacturer, there is a far smaller set of vehicles with whom a vehicle can be identical regarding its dimensions in comparison to the set of all vehicles from all manufacturers. Thus, we do not recommend to use only the discretization approach.

5.4 Summary of Evaluation

The obtained results show that, even without other tracking mechanisms, an attacker can track a vehicle with high probability using just a small set of constant volatile data, even though the vehicle performed a pseudonym change. This shows that the presence of volatile constant data is able to render PSC changes useless, as an attacker can re-identify vehicles using this data after the pseudonym change. Combining this attack with further tracking mechanisms, e.g., from [28], promises to achieve even higher tracking probabilities. Thus, the mechanisms for avoiding volatile constant data in VANET messages suggested in Section 3 should be used to limit the trackability of vehicles.

6 Conclusion and Future Work

With upcoming deployment privacy aspects of VANETs have gained increased attention, especially in Europe. Therefore, we studied the influence of information currently present in ETSI ITS and WAVE standard messages on various protocol layers on proposed privacy protecting pseudonym usage strategies.

We find that the main requirement of pseudonym change strategies, the cooperation of multiple indistinguishable vehicles, is unlikely to be found in practice with current standards being in use. This is caused by massive presence of individualizing content, which we call volatile constant data, within current VANET messages. It can be used by an attacker to easily distinguish vehicles independently of changing vehicle identities. Multiple suggestions have been made to improve this situation, which require to adjust corresponding standards within ETSI ITS and WAVE frameworks.

A metric called vehicle uniqueness is introduced to measure the chance of a vehicle to find proper cooperation partners for a secure, i.e., privacy ensuring, pseudonym change. Our evaluation shows that the currently standardized uncoordinated pseudonym switching yields very low probabilities for secure pseudonym changes. Limiting the amount of volatile constant data significantly increases the level of vehicular privacy within VANETs by decreasing vehicle uniqueness.

Future work can study the influence of the suggested privacy enhancement mechanisms on VANET applications, e.g., driver assistance systems. Such systems will probably need adjusted parameters to work based on more discretized, i.e. less accurate, input data sets.

References

1. Intelligent Transport Systems (ITS); Security; Security Services and Architecture (Sept 2010), v1.1.1
2. Memorandum of Understanding for OEMs within the CAR 2 CAR Communication Consortium on Deployment Strategy for cooperative ITS in Europe (June 2011), v 4.0102
3. IEEE Standard for Wireless Access in Vehicular Environments - Security Services for Applications and Management Messages (Apr 2013), 1609.2-2013
4. Intelligent Transport Systems (ITS); Facilities layer function; Facility Position and time management (2013), v0.0.2
5. Intelligent Transport Systems (ITS); Security; Security header and certificate formats (Apr 2013), v1.1.1
6. C2C-CC Basic System Standards Profile (Jan 2014)
7. Intelligent Transport Systems (ITS); Users and applications requirements; Part 2: Applications and facilities layer common data dictionary (Sept 2014), v1.2.1
8. Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Cooperative Awareness Basic Service (Nov 2014), v1.3.2
9. Intelligent Transport Systems (ITS); Security; Security header and certificate formats (June 2015), v1.2.1
10. Ammoun, S., Nashashibi, F.: Real time trajectory prediction for collision risk estimation between vehicles. In: IEEE 5th International Conference on Intelligent Computer Communication and Processing (2009)
11. Bissmeyer, N., Stubing, H., Schoch, E., Gotz, S., Stotz, J.P., B. Lonc, B.: A Generic Public Key Infrastructure for Securing Car-to-X Communication. In: 18th ITS World Congress (2011)
12. Bittl, S., Gonzalez, A.A.: Privacy Issues and Pitfalls in VANET Standards. In: 1st International Conference on Vehicular Intelligent Transport Systems. pp. 144 – 151 (May 2015)
13. Bittl, S., Gonzalez, A.A., Heidrich, W.: Performance Comparision of Encoding Schemes for ETSI ITS C2X Communication Systems. In: Third International Conference on Advances in Vehicular Systems, Technologies and Applications. pp. 58–63 (June 2014)
14. Burmester, M., Magkos, E., Chrissikopoulos, V.: Strengthening privacy protection in vanets. In: Proceedings of 8th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications. pp. 508 – 513 (2008)

15. Buttyan, L., Holczer, T., Weimerskirch, A., Whyte, W.: SLOW: A Practical pseudonym changing scheme for location privacy in VANETs. In: IEEE Vehicular Networking Conference. pp. 1–8 (Oct 2009)
16. Eichler, S.: Strategies for Pseudonym Changes in Vehicular Ad Hoc Networks depending on Node Mobility. In: IEEE Intelligent Vehicles Symposium (June 2007)
17. Gerlach, M., Güttler, F.: Privacy in VANETs using Changing Pseudonyms - Ideal and Real. In: 65th IEEE Vehicular Technology Conference. pp. 2521–2525 (Apr 2007)
18. Hoh, B., Gruteser, M., Xiong, H., Alrabady, A.: Achieving Guaranteed Anonymity in GPS Traces via Uncertainty-Aware Path Cloaking. IEEE Transactions on Mobile Computing 9(8), 1089 – 1107 (Aug 2010)
19. Houenou, A., Bonnifait, P., Cherfaoui, V., Yao, W.: Vehicle Trajectory Prediction based on Motion Model and Maneuver Recognition. In: IEEE International Conference on Intelligent Robots and Systems. pp. 4363–4369 (Nov 2013)
20. J. Harding et. al.: Vehicle-to-Vehicle Communications: Readiness of V2V Technology for Application. Tech. Rep. DOT HS 812 014, Washington, DC: National Highway Traffic Safety Administration (Aug 2014)
21. Kraftfahrt-Bundesamt: Neuzulassungen von Personenkraftwagen im August 2014 nach Marken und Modellreihen. online (2014), available `http://www.kba.de/DE/Statistik/Fahrzeuge/Neuzulassungen/MonatlicheNeuzulassungen/monatl_neuzulassungen_node.html`
22. Lefévre, S., Petit, J., Bajcsy, R., Laugier, C., Kargl, F.: Impact of V2X Privacy Strategies on Intersection Collision Avoidance Systems. In: IEEE Vehicular Networking Conference. pp. 71 – 78 (Dec 2013)
23. Nowdehi, N., Olovsson, T.: Experiences from Implementing the ETSI ITS SecuredMessage Service. In: IEEE Intelligent Vehicles Symposium. pp. 1055–1060 (2014)
24. Petit, J., Schaub, F., Feiri, M., Kargl, F.: Pseudonym Schemes in Vehicular Networks: A Survey. IEEE Communication Surveys & Tutorials 17(1), 228 – 255 (2015)
25. Scheuer, F., Plößl, K., Federrath, H.: Preventing Profile Generation in Vehicular Networks. In: IEEE WiMob. pp. 520–525 (2008)
26. Schütze, T.: Automotive Security: Cryptography for Car2X Communication. In: Embedded World Conference. pp. 1–16 (Mar 2011)
27. Stübing, H.: Multilayered Security and Privacy Protection in Car-to-X Networks. Springer Vieweg, 1st edn. (2013)
28. Tomandl, A., Scheuer, F., Federrath, H.: Simulation-based Evaluation of Techniques for Privacy Protection in VANETs. In: IEEE 8th International Conference on Wireless and Mobile Computing, Networking and Communications. pp. 165–172 (2012)
29. Wiedersheim, B., Ma, Z., Kargl, F., Papadimitratos, P.: Privacy in Inter-Vehicular Networks: Why simple pseudonym change is not enough. In: Seventh International Conference on Wireless On-demand Network Systems and Services. pp. 176–183 (Feb 2010)