Abstract—Vehicular ad hoc networks (VANETs) and vehicular communications are considered a milestone in improving the safety, efficiency and convenience in transportation.

Vehicular ad hoc networks and many vehicular applications rely on periodic broadcast of the vehicles’ location. For example, the location of vehicles can be used for detecting and avoiding collisions or geographical routing of data to disseminate warning messages. At the same time, this information can be used to track the users’ whereabouts. Protecting the location privacy of the users of VANETs is important, because lack of privacy may hinder the broad acceptance of this technology.

Frequently changing pseudonyms are commonly accepted as a solution to protect the privacy in VANETs. In this paper, we discuss their effectiveness and different methods to change pseudonyms. We introduce the context mix model that can be used to describe pseudonym change algorithms. Further, we assess in which situations, i.e. mix contexts, a pseudonym change is most effective and improves the privacy in vehicular environments.

I. INTRODUCTION

Vehicular ad hoc networks (VANETs) and vehicular communications will significantly improve the safety, efficiency and convenience in transportation.

By means of ad hoc vehicle-to-infrastructure and vehicle-to-vehicle communication the driver can be warned about critical traffic situations earlier and more precise. Further, cooperative safety applications between vehicles can be realized where vehicles exchange position data to prevent crashes. Last not least, convenience applications, such as point of interest notification and in-car access to the Internet can cost-efficiently be delivered by this technology. With the increasing number of vehicles to be equipped with a vehicular communication system this will probably result in the largest ad hoc network ever deployed. For vehicular ad hoc networks, privacy has been identified to be profoundly important (e.g., [1], [2], [3], [4]). When it comes to privacy in VANETs, many authors propose changing pseudonyms, yet do not discuss or evaluate how these pseudonyms can be changed to provide the best level of privacy.

In this paper, we briefly review attacks on changing pseudonyms. Further, we propose an approach that optimizes the use of pseudonyms while improving the overall privacy of the system, called context mix and show its superiority over arbitrarily changing pseudonyms by simulation.

This paper is organized as follows: in Sec. II, we discuss related work. We then discuss how attacks on the privacy can be carried out by linking changing pseudonyms in Sec. III. In Sec. IV, we introduce the context mix model and criteria for good situations to change pseudonyms. We evaluate the proposed situations – called mix contexts – by means of simulations and describe our findings in Sec. V. We conclude this paper and point out future work in Sec. VI.

II. RELATED WORK

In [5], Golle et al. propose using self-assigned digital pseudonyms. They further sketch three different measures to increase anonymity with changing pseudonyms that are similar to our mix contexts.

In [4], Dötzer discusses potential implications of missing privacy and proposes to use centrally assigned digital pseudonyms. He also proposes an architecture for pseudonym generation and registration for revocation purposes.

Huang et al. propose an approach based on dynamically created clusters [6]. The cluster-head works as a mix and all communication to the cluster-head is encrypted. The work focuses on non-safety applications. The authors further examine the use of a silence period to increase the effect of a pseudonym change. A common criticism is that clustering may be non-trivial to achieve in the dynamic vehicular environment.

In [7], Choi et al. propose a system to balance audibility and privacy in vehicular communications based on symmetric cryptographic primitives and two types of pseudonyms, namely long term and short term. They also propose to use changing short term pseudonyms for privacy protection.

Location privacy for pervasive computing has been treated in detail by Beresford in [8], [9]. He defines location privacy as “(...) the ability to prevent others from learning one’s current or past location” [8]. Beresford provides a detailed analytical analysis of the effectiveness of pseudonym change in mix zones. The approach is suitable for applications, which only need location information in dedicated and fixed application zones. The zones where an application does not need (to send) location information are called mix zones, because here, applications can change their pseudonyms.

The variable quality approach, where a centralized location server adjusts the accuracy of the given location such that the given area holds enough nodes to provide anonymity has been proposed by Gruteser in [10] and refined by Beresford in [9]. In [11], Gruteser and Huh discuss the anonymity of periodic location samples and describe one possible attack on linking messages based on Kalman filters.

III. ATTACKS ON PSEUDONYM CHANGES

Simply changing pseudonyms in arbitrary intervals and in arbitrary situations – i.e., contexts – wastes pseudonyms (and
hence resources for storing or calculating them) as Beresford already pointed out in [8]. E.g., for vehicular scenarios, two vehicles are clearly distinguishable by their direction. Hence, even if changing their pseudonym at the same time and being near to each other, they can be identified, and the pseudonyms are wasted.

A. Attack parameters

The actual cost of linking pseudonyms – if feasible at all – depends on the available information to the attacker and the algorithms and possible optimizations to link messages. This in turn depends on the following parameters:

- **Algorithms and rules used by the attacker** that can be used for linking messages. This includes simple approaches like matching the direction of nodes to multi target tracking as described in [11]. This also takes into account which knowledge an attacker uses to infer the linking.
- **Density and distribution of receivers.** The denser the grid of receivers is, the higher will be the probability of an attacker overhearing a pseudonym change. In our work, we assume a global attacker.
- **Beacon frequency, type, and accuracy of information sent.** A higher beacon frequency will probably make tracking easier because a higher beacon frequency narrows the area for matching two messages. Further, if any identifying (non-volatile) data is sent in the beacons in addition to the pseudonyms, this can be used by the attacker to link two pseudonyms as well.

B. Potential algorithms and rules

We identify three major directions that an attacker could follow to link pseudonyms:

- **Attacks based on non-volatile data**, where additional data that does not change (such as unencrypted higher layer identifiers, or the radio fingerprint of a unit) are used to infer a connection between two messages.
- **Protocol based attacks**, where knowledge about the protocol (e.g., a vehicle always sends in a particular time-slot, independent of its pseudonym) is used to link messages.
- **Attacks based on physical parameters and constraints**, where knowledge about, e.g., the estimated distance travelled and the last position is used to infer the current position, and hence to link two messages as belonging to the same node. This attack has been described as simple tracking by Huang et al. [6].

Attacks based on non-volatile data will probably be the cheapest. Their effectiveness is high, if the non-volatile data can be used to distinguish between the different nodes that change their pseudonym in a certain area. A commonly cited attack is the use of the radio fingerprint of the communication system.

Different algorithms for tracking pseudonym changes have been proposed based on physical parameters, such as finding a maximum match in a bipartite graph using pre-established data on the movement of nodes in a mix zone (Beresford in [9]) or using Kalman-filter based techniques (Gruteser and Hoh in [11]). Both techniques are based on maximizing the probability of two pseudonyms belonging to the same node, based on physical parameters.

Clearly, all attacks aim at finding properties of a node that make it identifiable and recognizable. For our evaluations, we use two rather simple attackers that will be described in Sec. V.

IV. CONTEXT MIXES – INCREASING THE PRIVACY IN VANETS

We propose to include the use of context information (such as the number of neighbors, their direction and speed) for initiating a pseudonym change. Like this, nodes cooperatively identify good opportunities to blend in a number of vehicles and hence increase their anonymity. Following the terms mix-zones (Beresford) and mix-nets (Chaum) we call these situations mix-contexts.

Fig. 1 depicts a general state diagram of a pseudonym change algorithm. The minimal stable time may be configured to account for the application requirement of a stable communication session. After the stable time finishes, the node waits for the trigger to change its pseudonym, checks if the change has been successful and then enters the next period of stable pseudonym to run through the process again.

After initialization, the system enters the pseudonym cycle and waits for expiry of the stable time interval. Under certain circumstances, a pseudonym change may be sensible before the stable time is over; in this case the stable time is overridden. The system is then ready to change its pseudonym, and in this state permanently assesses its context (i.e., neighborhood information) in search for a mix context that suffices the target level of anonymity. If this mix context is eventually found, a new pseudonym is retrieved and set. Simply put, the target level of anonymity can be a certain number of nodes with similar direction within a certain range. After changing the pseudonym, the system assesses whether the change was successful (i.e., if enough similar nodes changed their pseudonym at the same time) or not in order to start the whole process again, or try to change the pseudonym again, respectively.
A. Pseudonym change triggers - mix contexts

Dey defines context as "(...) any information that can be used to characterize the situation of an entity (...)" [12]. Using this definition, a mix context is defined as any situation that provides sufficient anonymity with respect to an attacker to change a pseudonym.

Depending on the desired level of protection, this may simply be the number of nodes in the neighborhood irrespective of their properties, or the nodes with similar properties, such that they would be indistinguishable for an attacker. A pseudonym change algorithm using in mix contexts is a context mix. A context mix provides unlinkability between pseudonyms after a change.

A mix context shall provide sufficient anonymity to a node changing its pseudonym. This requires that the neighborhood of the node and the general situation must be such that the entropy of the situation after the change is sufficiently high. Hence, a node must permanently assess its context according to the expected entropy if it changes its pseudonym. The expected entropy also depends on the attacker; this implies that every node may need to implement a reference attacker to estimate its level of privacy.

In addition to the expected entropy, i.e., the anonymity of the change, the potential impact of the pseudonym change is important. Situations that allow a direct mapping of the pseudonym to the user, for example by restricted space identification, may require a pseudonym change shortly before and after this situation in order to limit the amount of available information on the traces for the identified user.

Currently, we define the availability of more than $N$ nodes in a defined area as mix context. In addition to simply changing the pseudonym in the right context, we define a minimal stable time where the node is supposed not to change its pseudonyms. This is important in order to prevent frequently terminated connections, and it bounds the number pseudonyms used per node.

V. ANONYMITY SIMULATIONS

For our simulations we used JIST/SWANS [13] and the vehicular mobility model provided with STRAW [14]. The simulation engine is written in pure Java and runs within a standard Java virtual machine, by embedding simulation time semantics during execution at byte-code level. The Street Random Waypoint (STRAW) mobility model allows to simulate maps of large real world cites and offers advanced vehicle behavior together with simplified traffic control mechanisms.

A. Simulation parameters

The following parameters were changed in the simulation runs below:

- traffic density,
- pseudonym change algorithm, and
- attacker model.

The traffic density defines how many vehicles could be found on one kilometer street length at specific point in time. The Forschungsgesellschaft für Straßen und Verkehrswesen (FGSV – research agency for roads and transport) classifies five different traffic density ranges as follows [15]:

- $< 16 \text{ vehicles per km}$ low traffic density
- $16 - 23 \text{ vehicles per km}$ medium traffic density
- $24 - 31 \text{ vehicles per km}$ high traffic density
- $32 - 45 \text{ vehicles per km}$ very high traffic density
- $> 45 \text{ vehicles per km}$ overload

The simulation uses a map of a real urban area. It contains different street types with assigned speed limits ranging from 11 meter per second to 19 meter per second. The total street length is about 16 km. All segments have two directions and at least two lanes. One way streets are currently not supported. In line with the traffic densities defined above we simulate low traffic density with about 6 nodes per kilometer (100 nodes on map) and about 13 nodes per kilometer (200 nodes on map) as well as medium traffic density with about 19 nodes per kilometer (300 nodes on map).

The two pseudonym change algorithms simulated for this paper are: random pseudonym change and context mix pseudonym change. Both algorithms keep the pseudonym stable for a minimum stable time of one minute. This value has been chosen because it represents a reasonable value for position based routing [16], [17].

The random algorithm changes decides if it changes its pseudonym for every beacon it sends (except in the minimum stable time). It can be configured with a probability threshold that is compared with a randomly generated value every beacon interval. The pseudonym is changed if the random value is below the probability threshold. As the reference algorithm to compare with the context mix concept, we manually adjusted the probability threshold to find the best results in several simulation runs not described here. Arguably, this algorithm is better than a fixed time interval for pseudonym change since with the fixed interval a pattern exists that can easily be followed by an attacker.

The context mix algorithm only changes pseudonym if in the preset mix context. A mix context at best represents all information that an attacker may use to link pseudonyms. In our case, the mix context is limited to the information our attackers are using and which are provided by each vehicle. This comprises the vehicle’s last position and the pseudonym used\(^1\). We currently do not incorporate velocity, heading, acceleration or other helping context information to separate vehicles in our pseudonym change and attacker algorithms. If the node neighborhood includes $N$ vehicles at a distance smaller than the minimal distance for a pseudonym change, it changes its pseudonym. The minimal distance for a pseudonym change is set to 4.25 meters according to an average lane width of 3.5 meters and an average position reporting error of 0.75 meters.

We implemented two attacker models. A simple attacker and a multi target tracking attacker, which are two different stateless attackers. The attackers decide on the fly if they can

\(^1\) In the future vehicular system, the pseudonym will comprise the MAC and IP address of the vehicle and its certificate for network access control.
trace a vehicle. Simple tracking expects the vehicle to send a message in a fixed area around where the previous message was sent. It fails if there are several vehicles in that region. Multi target tracking can link the pseudonym if it detects that only one vehicle in a set of suspects within the expected area has changed the pseudonym. Multi-target tracking fails if the anonymity set size is greater than one after all detectable, innocent suspects are excluded. Both algorithms use information like the maximum speed and position accuracy, which are not included in the mix context definition because these values are provided by the system, not by the vehicle.

The location update cycle time is the frequency of location update broadcasts done by each vehicle in the simulation. In this work we focus on a fixed location update rate of $1\,Hz$. This value has been chosen because it represents a reasonable update rate for e.g. position based routing and simple warning applications.

B. Simulation results

Fig. 2 shows the results of a simulation run with 100 vehicles sending a beacon every second under the multi target tracking attacker. The minimum stable time was set to a minute. The values on the y-axis represent the number of pseudonyms that could be linked at after a certain time ($T$). Simulation time was 30 minutes.

The figure shows that less vehicles could be tracked if they change their pseudonym in mix contexts. It also shows that a large portion of vehicles can still be tracked even though they change their pseudonyms. We are currently exploring this behavior in additional simulations and expect to improve the mix context algorithm by adjusting context parameters. In addition to the better privacy provision by the mix context algorithm, the overall number of pseudonyms used per vehicle is smaller, because a pseudonym is only changed when it may really be worth it.

The figure also shows how long a multi-target tracking attacker is able to follow a vehicles trace for both pseudonym change algorithms. The $x$-axis shows the average time, including a minimal stable time of 1 minute that we used in the simulation. The average tracking time is influenced by the traffic density and the pseudonym change algorithm. The following observations can be made:

1) A higher node density leads to shorter pseudonym usage times for both algorithms.
2) The node density has no significant influence on the performance difference between both algorithms.
3) Context mix algorithms successfully change pseudonym roughly 2.4 times faster than random pseudonym change algorithms.

Shorter tracking times lead to better privacy. After the minimal stable time passed it is desirable that a node successfully change its pseudonym as fast as possible.

Fig. 3 shows the worst case for privacy is the low traffic scenario, for both algorithms and the different traffic densities. The random pseudonym change algorithm is only able to reach pseudonymity for 21% (see Fig. 4) of all vehicles in the first minute after the minimal stable time has passed. There is also a significantly high number of vehicles that are not able to successfully change their pseudonym within 5 minutes. The context mix approach performs better. 81% of vehicles can change their pseudonym within the first 5 minutes unlinkable to our attacker. The majority of these nodes is able to change within the first minute.

The results for the higher density vehicle simulations are slightly different, as depicted in Fig. 5. The context mix approach successfully changes pseudonyms for 91% and 96% in the first 5 minutes. Up to 78% of pseudonyms already could be successfully changed in the first minute at high traffic density. Under the same conditions, the random pseudonym change algorithm just changes 37% in the first minute.

From these results we can see that mix contexts algorithms perform well at high traffic and in comparison to random algorithms especially in low traffic scenarios.
The reader may notice that we concentrated on the rather lower end of traffic densities. Since VANET technology will not have full market penetration from the first day low traffic scenarios are interesting to measure algorithm performance. Currently we are further investigating the influence of market penetration (including non-VANET enabled vehicles in our simulations) on the performance of mix contexts pseudonym changes. Higher densities will probably yield better performance of both algorithms.

VI. CONCLUSION AND FUTURE WORK

The simulations show an improvement in the achieved level of privacy for this approach. Currently, more simulations to verify and refine these results are carried out and will be included in subsequent work. Another advantage of this approach is the more efficient use of pseudonyms due to only changing them when it improves the privacy. On the other hand, a couple of issues have to be taken in mind: First, our simulations showed that the minimum stable time affects changing pseudonyms, because the probability to meet a node changing its pseudonym decreases. Therefore we introduced a change ready flag that is broadcast by a node where the minimal stable time expired. Thus when two nodes with this flag set meet, they the probability that they will change their pseudonym at the same time increases. The use of this flag will be examined in more detail in future work.

Second, if different nodes take different context information into account, they will change their pseudonyms in different situations. In addition, the more context information is considered, the fewer situations will occur where a node changes its pseudonym. Thus, it may be important that pseudonym change algorithms are the same for all nodes in the network.

Third, the parameters for the algorithms need to be refined in order to optimize the privacy provisions. In particular, minimum stable time will need to be adjusted to realistic values and its impact examined.

Finally, the applicability of the algorithm in real life scenarios still has to be proved. This includes estimating a sensible minimal stable time, including data about when the vehicle is started, and the like.

In a nutshell, mix contexts provide an improvement of the anonymity in vehicular ad hoc networks over randomly changing the pseudonyms in certain intervals. The complexity of the algorithm is low, as vehicles do not require explicit group formation to change pseudonyms. Looking at the simulation results, however, reveals that the amount of tracking that a global passive attacker can achieve is still significant.

ACKNOWLEDGEMENTS

This work has been carried out in the ”Network on Wheels” [18] project supported by the German Ministry for Education and Research under Contract No. 01AK064F.

REFERENCES