

Full Paper: Assessing and Improving Privacy in VANETs

Matthias Gerlach, matthias.gerlach@fokus.fraunhofer.de

Abstract—Vehicular ad hoc networks (VANETs) and many vehicular applications rely on periodic broadcast of location information of cars. At the same time, this information can be used to track the users whereabouts. Protecting the privacy of the users using VANETs is important, because lack of privacy may hinder the broad acceptance of this technology.

We carry out an extensive risk analysis for VANETs, where we identify attacks threatening the privacy of users based on a simple cost estimation. We explicitly include existing technologies such as cellular phones in the analysis and argue that using VANET technology makes attacks on privacy less expensive. Further we identify – and rectify, where those measures have already been proposed – countermeasures against the most common attacks.

As changing pseudonyms addresses many of the privacy problems that are discussed in this paper, their effectiveness is discussed as well. Finally, we propose and discuss a novel approach to changing pseudonyms, called *mix contexts* to improve the privacy of users in VANETs. We conclude by outlining future research directions in the field.

I. INTRODUCTION

For vehicular ad hoc networks, privacy has been identified to be profoundly important (e.g. [1], [2], [3], [4]). Despite the identified importance of privacy, there is yet no methodological risk analysis for VANETs. This often leads to only a vague idea on what is at stake and also to vague discussions. Therefore, one contribution of this paper is a description of likely attacks on privacy when using a VANET based on a thorough risk analysis. As a major contribution, the risk to privacy is quantified and possible countermeasures are discussed. We further propose *mix contexts* as a means to change pseudonyms in a way that makes it hard for an attacker to track a vehicle.

A. Approach

The risk analysis is carried out using attack trees as have been introduced in [5] and extends the work by Aijaz et al. in [2]. For creating the attack trees, we developed AttackML, an XML-based language to create and assess attack trees. The language helped in marking up the trees, assessing and modifying them, and allowed for textual (XML) and graphical viewing and editing in the Eclipse IDE¹.

The initial assumption of the risk analysis is that every node uses only one unique and permanent pseudonym (identifier) for communication. Further, we assume that the vehicle periodically emits messages including its position, a time stamp and its pseudonym. For many applications, such as position based routing and many safety applications, this is a valid

assumption. For comparison, we also included sub-attacks that are not based on the vehicular technology itself, like cellular phone tracking, and video surveillance.

The analysis shows that changing pseudonyms will improve the level of privacy significantly, by preventing most of the attacks described below. Yet, arbitrarily changing pseudonyms will not help, therefore we will propose and discuss *mix contexts* in Section VI.

B. Overview

In Section II, related work is discussed. If an attacker's goal is to obtain privacy sensitive information of a user, two major attack goals have to be accomplished. First, the attacker must *track the location* of a vehicle. He yields a list of pseudonyms and locations. This attack will be described in detail in Section III. Second, in order to use the location information, the attacker has to *link the pseudonym to the user*; we will look at this attack in Section IV. Section V discusses attacks to link changing pseudonyms. Consequently, we propose and discuss an improved approach to privacy in VANETs using *mix contexts* in Section VI. In Section VII, some real world attacks against the privacy are outlined. We conclude with a discussion of probable attacks, attackers and possible countermeasures in Section VIII and outline future research directions in Section VIII.

II. RELATED WORK

In [4], Doetzer discusses potential implications of missing privacy and proposes to use centrally assigned digital pseudonyms. These could be traced back to a single vehicle only by a trusted authority.

In [6], Golle et al. propose using self-assigned digital pseudonyms. They further sketch three different measures to increase anonymity while changing pseudonyms: first, synchronized pseudonym change, second, introducing gaps in data reporting after pseudonym change, and third, changing pseudonyms, when nodes are near to each other. We call these situations *mix-contexts* and extend the idea in Section VI.

Sampigethaya et al. propose an approach based on dynamically created clusters, where the cluster-head functions as a mix and all communication to the cluster-head is encrypted [7]. Similar to Golle et al., the authors propose silence periods to make tracking a pseudonym harder.

Each of the above contributions contains a high level threat analysis. These risk analyses are often challenged arguing that there are already many technologies out there that threaten the privacy of the drivers. We support the reasoning in favor of

¹www.eclipse.org

privacy measures for VANET communication backed up with the risk analysis below.

This paper is the first extensive risk analysis for privacy threats in VANETs. For wireless networks Gruteser et al. proposed a methodology for assessing privacy in wireless network in [8]. While their approach is similar to ours and also identifies tracking *and* identification as the two aspects of a successful attacks, the authors focus on location-based services rather than applications in VANETs. Our work goes into more detail than the authors' in that a cost estimation for different attacks is provided. In [9] Gruteser and Huh discuss the anonymity of periodic location samples and describe one possible attack on linking messages based on Kalman filters as mentioned in Section V.

Location privacy for pervasive computing has been treated in detail by Beresford in [10], [11]. He defines location privacy as "(..) the ability to prevent others from learning one's current or past location" [10]. 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 *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.

III. LOCATION TRACKING

This is the attack described most often for vehicular ad hoc networks. Note that we argue that this attack alone does not violate the privacy of a user. This is because the pseudonym hides the real identity of the user and thus protects his privacy. As the proceeding of an attacker is the same for almost any technology that can be used for tracking a node, we identify a general attack method and later parametrize it for the different sub-attacks before discussing the actual attacks.

A. Attack Method

The objective of the location tracking attack is to create a database allowing for queries such as "where was node X on Sunday around eleven" using a setup as depicted in Figure 1. In this section, we focus on collecting tuples containing an identifier (the pseudonym), the time and the location of the node.

In order to be able to track a vehicle, the following steps are necessary:

- 1) *Build a grid of receivers.* Receivers can be represented by any technology such as RFID², cameras, cellular phones, or even people.
- 2) *Connect these receivers to the central database.* This may be as simple as using an existing broadband connection or as expensive as using a satellite connection.
- 3) *Store the data.* Solutions for storing data range from a simple hard-disk to storage area networks and expensive large storage servers.

²Radio Frequency Identification.

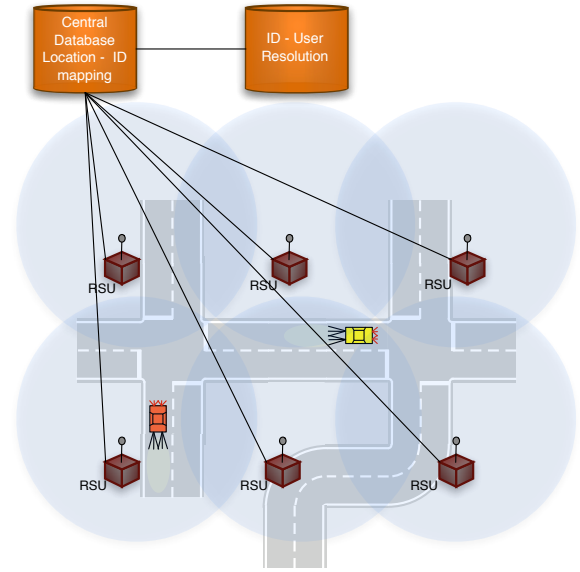


Fig. 1. Tracking the Location of Nodes using a Grid of Receivers - General Architecture.

- 4) *Process the data.* Depending on the amount of data and the types of tasks either a simple ordinary PC can do the task, or rather expensive high performance workstations must be used.

As all of these attacks have to be carried out (they are represented by an AND connection in the attack tree), their cost add up. In Sections III-C to III-E, we estimate the cost for the different attack scenarios by parameterizing the above attack with the values defined in the next section. An overview of the estimated cost for different aspects of the attack can be found in Table I. In order to compare the cost of an attack we only give indicative values.

B. Attack Parameters

An attacker can have different goals with respect to the data he wants to collect. Basically, the number of nodes and their properties to be tracked are relevant, as well as the area under surveillance. With respect to these goals, we classify the possible attack into four dimensions:

- All nodes, everywhere (See Section III-C).
- Some nodes, everywhere (See Section III-D).
- All nodes, some place (See Section III-E).
- Some nodes, some place (See Section III-E).

All nodes in this classification refer to all existing nodes having a transceiver, and *everywhere* would describe a ubiquitous, global grid of surveillance. *Some nodes*, on the other hand describe a sufficiently small number of nodes having a certain property. The number of nodes is interpreted as the first attack parameter.

The second attack parameter is the level of coverage of the attacker. Here, *some place* describes one (or several) important places for the attacker to monitor.

Item	Fix Cost ^a	Cost p. Month ^a	Remarks
Road Side Unit	+	+	
Camera for Video Surveillance	++	+	Camera must support license plate recognition and have line of sight.
Receiver for cellular phones	++	+	
Human		+++	
RFID Reader	++	+	
Broadband connection	+	+	May already be available.
Cellular connection	+	+	
Storage Space	+++ p. TByte ^b	+++	Depending on storage technology.
Processing power	Depending on the amount of data and the complexity of the queries.		
Access to database of phone provider	Only if legitimized.		
Access to database of camera surveillance / toll collect company	Only if legitimized.		
Install malware on vehicle	Need access to vehicle application platform.		

TABLE I
ESTIMATED COST FOR PARTS OF THE ATTACKS.

^a+ – order of 10 EUR, ++ – order of 100 EUR, +++ – order of 1000 EUR.

^bFor example, a 9 Terabyte ProLiant Data Protection Storage Server is available for about 33 K EUR.

A third parameter is how long traces are stored. For this analysis we take a month as basis for the cost estimation. Longer attacks are simply assumed to increase the cost linearly.

C. All nodes, everywhere

In this ubiquitous surveillance scenario, any vehicle can be located at any place.

With cellular networks, or – in Great Britain a reality – a dense grid of surveillance cameras, there are readily available techniques for tracking a car. Using cellular networks, the position of a user can be tracked with cell-size-accuracy (a cell is between 30m and 1000m in radius), while cameras monitor number plates at fixed locations only. Even though these techniques can be used today to track vehicles, access to the data depends on the appropriate legitimation to access or the capability to build an own such infrastructure.

Building own grids of transceivers for cellular technologies on the other hand is more expensive than using road side units. An additional advantage of a wireless communication technology over, e.g. near field communication (such as RFID) or cameras, is that no direct line of sight or additional processing is necessary to obtain the pseudonym. Last but not least, installing Road Side Units (RSUs) may have additional benefit to the installing parties (e.g. for drive through payment, infrastructure based safety applications, or point of interest notification).

Simple calculations show that even though still the cheapest choice over established technologies (cellular, cameras), tracking vehicles using a grid of RSUs in this scenario is costly, in particular because setting up and using the tracking infrastructure would be expensive (in particular in the order of some million Euros for using and maintaining the set up infrastructure.)

There are several ways to reduce the cost for potential attackers. First, similar to current activities in loyalty cards, where several companies use one provider, companies could

team up to share the cost of this attack. In particular, cost for installing and connecting receivers can be saved using existing broadband connections on existing sites and sharing receivers with other parties.

Despite the potential savings, we can state that in the medium term, the threat of being traceable in the *all nodes everywhere* attack will not be significant. In particular, the cost and effort to set up a grid of receivers and to collect and store the information is too high for the attack to be probable for a single stakeholder. With this respect it may be easier to use well-established techniques for tracking, such as cellular phones and grids of cameras. Note that both techniques require proper legitimation to use them.

On the other hand, it may be a business-model to sell information about when a user's whereabouts in the future, e.g. for advertisements. This model will be even more interesting, the easier it is to resolve the user. A permanent identifier for communication would be a perfect support for those kinds of systems. Note that in the long run, the major problem of an attacker will be to extract useful information rather than merely collecting data.

Assumed an attacker already has access to a grid of receivers and is able to build a database of locations the following measures can be thought of to make the task of location tracking harder for him:

- Change the identifiers such that they cannot be linked.
- Do not provide more accurate data than necessary.
- Do not provide more data than necessary.

D. Some nodes, everywhere

If a private investigator seeks to track the whereabouts of his target and wants to make his life easier using VANET technology, there are three possibilities once he knows the pseudonym to track: (1) by means of frequently sent information by the target node, (2) by means of the location service of the routing protocol (3) using an application that frequently reports the locations of a vehicle to the attacker. Most VANET

routing protocols use positions in order to better adapt to the dynamic topology, and these routing protocols provide a facility to resolve the location of a specified pseudonym [12].

The first attack's cost is basically the cost for building the infrastructure in the *All nodes, everywhere*-attack in Section III-C to track the target node everywhere. The cost for storage and processing will probably be negligible in this scenario.

The second option basically uses the existing network as the grid of receivers, and is significantly cheaper. The location service of the routing protocol can be used to track a selected node. In order to do this, the attacker has to obtain a transceiver and install an application that frequently requests the location of the sought node. The location service then returns the corresponding location. Therefore, the "grid of receiver" consists of a single receiver plus the receivers in every VANET node.

The cost of this attack is in the order of some hundred Euros and can be carried out with any legitimate transceiver. In addition, if a target node has no means to control neighboring nodes to report back its location, it cannot prevent an attacker from learning its location. This could make the attack even more efficient, as the attacked node does not know that it is being tracked.

The effectiveness of this attack for a single user is restricted by the functionality of the location service. If the scope of a location query is bounded by a number of hops, then a user can only be traced in that radius. In addition, this attack requires a densely connected network up to the target node.

Due to the maximum network load, it is improbable that this attack supports tracking many users at the same time, because location request are assumed to raise the network load.

We conclude that while this can be one of the most probable small scale attacks on the privacy of a user, it is unsuitable for large scale use, and very limited by the network connectivity and dynamics.

Finally, installing a piece of malware at the target vehicle that records and reports back the location of this vehicle may be a cheap option for an attacker. As the node itself records back its location, not even a receiver may be necessary, as long as the infrastructure for the malware to report back its data exists. As some of the applications in VANETs require at least sporadic infrastructure access, this may come for free as well. Note however, that installing an application on the vehicle may not be as easy as on a commodity PC today; yet, countermeasures should be taken to prevent the unauthorized installation of applications.

Possible countermeasures can be to:

- Artificially restrict the max number of hops for a *location query*, or
- Change the pseudonym frequently such that a location query fails, or
- Forbid frequent location queries by specification from the network side.
- Control the installation of applications on the vehicle to prevent malware.

E. All/Some nodes, some place

This attack is useful to observe locations with a particular significance. Imagine a criminal that wants to blackmail all people attending a certain conference (they parked on the conference hotel parking), or some establishment. He gets the list of pseudonyms and then needs to map the name to the pseudonym (see Section IV).

In order to observe a selected location, the attacker installs a receiver near that location. Therefore no *grid of receivers* is needed. Like this, the attacker can record the identifiers of all vehicles that broadcast to be at the monitored location.

Depending on where the receiver is located, *connecting the receiver* to the database is cheap or can be done using one of the technologies listed above, where the cost does not exceed 100 EUR.

Storing and processing the data of all passing-by vehicles can be managed by a state-of-the-art PC.

In addition to be cheap, this attack works by simply overhearing beacons of the vehicles, and is therefore hard to detect. The attacker obtains a list of pseudonyms who visited a particular location.

We conclude that this attack is a very probably attack, which will almost certainly be used to track the occurrence of vehicles at a certain location. With permanent pseudonyms, this attack can be used to re-recognize a node (but not the user) at a certain location.

As vehicles are assumed to permanently send beacons, there are no means to prevent an attacker from creating the list of pseudonyms except for switching off the transceiver. This in turn may not be desired because it may affect the network density in a negative way and even the safety of other cars.

IV. LINK VEHICLE PSEUDONYM TO USER

The previous attacks only yield the pseudonym in connection with the location and a time-stamp. For an attacker to use this information, he needs to obtain the name of the car holder or the driver³. Only then, the privacy of a user is affected. Under the assumption of a permanent pseudonym, an attacker has to carry out this attack only once and can afterwards use the information. We identify four possibilities to gain knowledge about the mapping of the pseudonym to the name of the holder:

- Use an existing database, or
- Restricted space identification, or
- Inference from external database information, or
- Ask the nodes.

The cost and feasibility of each of the approaches depends on the scale of the attack, i.e. for how many users the name-to-pseudonym mapping shall be obtained. We identify for what types of users this attack is feasible and what the attacker must do to carry out the attack.

³It is hard to tell from the attacker's point of view, who will sit in the car, therefore we will not make that difference either. We note however that this is imprecise and should be discussed further.

A. Use an Existing Database

The database referred to in this attack links pseudonyms to real world names. It can be seen as some kind of phone-book resolving pseudonyms to names (and address).

Note that currently, many solutions assume a database to link pseudonyms to names maintained by some authority in order to revoke faulty nodes and make people accountable for manipulating nodes (e.g. [4], [2], [1]). The cost of obtaining information will then depend on regulations and the actual implementation of the mapping. For the majority of attackers (such as private persons, most companies), using this database will not be feasible. However, any attacker creating an own such database using other means (such as the ones discussed below) may sell or give away this information.

Countermeasures can be to

- Not require such a mapping (for certain applications), or
- Establish appropriate legislation preventing the use of such data, or
- Distribute information such that several parties are required to federate to use the database.

B. Restricted Space Identification

Restricted space identification refers to the fact that with today's technology, a position can automatically be matched with additional information, a home address, a phone number, and the like. This attack has been introduced in [13].

Basically, any location where an *automated mapping* from a location to the name of the driver/holder makes possible restricted space identification.

As an example, in a rural area, where everyone has his own house with a parking (garage), such a mapping can be done in three steps:

- 1) *Obtain the location.* Typically, the rule “sleeps at home / car is at home overnight” should hold. Note that inferring this location may be as expensive as carrying out the *all nodes, everywhere* attack if it shall be carried out for an arbitrary node. In order to obtain this information for only specific areas (e.g. areas where wealthy people can be expected to live) the attacker can focus on only these areas.
- 2) Use existing database to *infer address from home location.* There are existing applications for this, such as those used for navigation system software.
- 3) Finally, *link the address to the name of the inhabitant.* The data to do this can be found in every phone book.

The example shows that in theory, such attacks are feasible. Note the importance of the one-to-one mapping from a location to a name (that can be detected automatically) for a successful attack. For many areas in the cities, this approach would fail due to the lack of such situations, mainly because people live in multifamily buildings where there is no personal parking space.

As mentioned above, obtaining the “one-to-one-mapping” location may be expensive if the attack shall be used in larger scale; hence simply following a car is cheaper for an attacker.

We argue that in order for this attack to be feasible for many victims, the attacker needs a grid of receivers. This makes the attack cost intensive and therefore rather improbable.

Possible countermeasures against this attack are:

- Avoid situations having a one-to-one mapping of name to location.
- Change the pseudonym shortly before and after visiting such locations, effectively making the mapping useless.

C. Inference from external database information

The objective of this attack is to infer the name-pseudonym mapping from data transmitted in the messages and external databases. In terms of statistical disclosure control, every such set of data is called *microdata*. The objective of the attacker is to link two sets of microdata: first, the information contained in the messages sent by a node (all these messages are linkable due to the permanent pseudonym); second, the information in external databases, which already contain a name and statistical data (such as a profile from a loyalty card programme).

As an example, suppose an attacker knows that a particular user (who just disclosed his name) drives a blue Ferrari, and there is just one blue Ferrari around at that time, it is easy to infer that the user belongs to the Ferrari and then link the name to the pseudonym of that Ferrari.

In order to carry out the attack successfully, an attacker must first create a user profile, and then link the profile to the vehicle.

There are already existing databases which contain profile information about customers. Therefore the first part of this attack will probably already be accomplished.

The feasibility of linking the profile information to the vehicle depends on the availability of sufficient data to link the two sets of microdata. This in turn depends on what data is collected in the external database (e.g. a loyalty card programme) and what non-volatile (i.e. identifying) information is sent out by the vehicle.

If a vehicle only sends out a pseudonym and volatile information such as speed, location, environmental data, it will probably be hard for an attacker to link vehicle identification and profile information. The linking can be made easier for the attacker by attaching identifying tokens to the user. This reduces the influence of the user on preventing the mapping. Examples are RFID tags, short range radio technologies (such as ZigBee or Bluetooth, if their id's are stored in the profile), parking tickets, and number plate information.

Possible countermeasures against statistical disclosure are:

- Reduce the accuracy of disclosed data (e.g. “blue vehicle” instead of “blue Ferrari”).
- Sparsely disclose data that can be used for statistical disclosure (such as any non-volatile identifiers), and hence minimize data in the external database (e.g. the mapping of the name of the user to the information that he drives a blue Ferrari).

- Change the pseudonym shortly before and after situations where profiling could be possible, effectively making the mapping useless.

D. Ask the nodes

The obvious – and cheapest – way of obtaining the name of the user is to simply ask him to send his name using the communication system. Then, the attacker automatically obtains the sought information. Example systems where name information are transmitted, are

- Payment methods, such as credit cards,
- Loyalty cards,
- Chat applications or any application using real names as pseudonyms.

In these scenarios, name information is given away to a specific – trusted⁴ – party for free. This party can then infer the name of the user wherever it wishes to even without the explicit consent of the user. Note that the information has to be encrypted to be protected against untrusted parties.

We note that this attack is very probable, because it is easy and the name of the user is used in some applications. Basically, it is similar to phishing attacks, which are very common nowadays. In addition, some applications ask the user once to enter his name and then decide autonomously when to disclose it. These applications may even more easily be tricked to disclose the name. Although the user has the possibility to opt-in by disclosing his name once, a permanent pseudonym leaves no room for opting out, and should be prevented, e.g. by changing pseudonyms.

V. LINK CHANGING PSEUDONYMS

From the above analysis, it becomes clear that the decision in previous work to use frequently changing pseudonyms is well founded. However, simply changing pseudonyms in arbitrary intervals and in arbitrary *contexts* is useless and wastes pseudonyms (and hence resources for storing or calculating them) as Beresford pointed out in prior work [10]. 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 Method

We assume that a vehicle has some sort of identity that is not used for communication; an attacker can make up a permanent identifier representing this identity. As they prevent a lot of attacks, we assume the use of frequently changing pseudonyms for the communication. These pseudonyms are independent such that an attacker cannot link any of them without further knowledge. For the period that a node just uses one pseudonym, linking messages is trivial.

The goal of an attacker is to link the pseudonyms to the identity of the car and consequently be able to link two pseudonyms. In order to do that, he needs to

- 1) *Set up an infrastructure for location tracking.* Similar to the location tracking attack described in Section III, the attacker needs to create the grid of receivers, link the receivers to a database, and store the data. The costs for these attacks remain the same as discussed above as we neglected processing cost there.
- 2) *Find (all) pseudonyms that belong to the same node.* This is basically step four in the attack method of location tracking (see Section III). The attacker now has to invest some effort in linking the messages from different pseudonyms. The effort and the feasibility of this depends on the attack parameters described in the next section.

B. 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* that can be used for linking messages. This includes simple approaches like matching the direction of nodes to multi target tracking as described in [9]. 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.
- *Beacon frequency and 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.

C. 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 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.

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.

Different algorithms 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

⁴I.e. the user trusts this party to treat his information confidential.

in a mix zone (Beresford in [11]) or using Kalman-filter based techniques (Gruteser and Hoh in [9]). Both techniques are based on maximizing the probability of two pseudonyms belonging to the same node based on physical parameters. A detailed quantitative discussion and comparison of the different attacks is left to future work.

VI. MIX CONTEXTS - INCREASING THE PRIVACY IN VANETS

Looking at the different attacks in Section V, however, it becomes clear that just changing a pseudonym is not sufficient. This fact can be confirmed based on the simulations and analyzes carried out by Sampigethaya et al. in [7]: the authors state, that under “correlation tracking”, an attack that uses physical parameters and constraints, changing the pseudonyms at arbitrary intervals yields an anonymity set below 2, and hence no anonymity at all [7, Figure 6].

Therefore, we propose to include the use of context information 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 this approach *mix-contexts*.

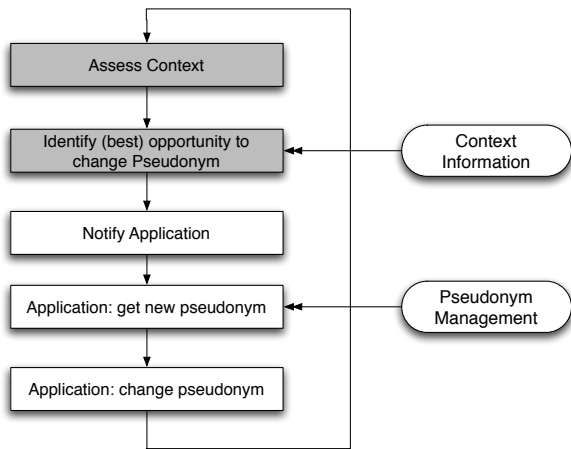


Fig. 2. General Approach to Realize Mix Contexts.

A. Description

Following Deys definition of a context as “(...) any information that can be used to characterize the situation of an entity (...)” [14], a mix context describes the available information relevant for taking a decision to change a pseudonym. Depending on the desired 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.

The general process for determining the mix context is depicted in Figure 2 where the main aspects, *assess context*, and *identify best opportunity to change a pseudonym* are highlighted. For assessing the context, each node will calculate

a value indicating the level of anonymity. A candidate measure is entropy, because one can express the level of privacy well in one value; as Serjantov et al. pointed out in [15], entropy can measure the level of uncertainty for an attacker in a specific situation to track a node, and hence provide a good estimate of the current anonymity⁵. To identify the best opportunity to change, a threshold for the minimum entropy has to be defined either by the user or by an application where a pseudonym change is triggered.

B. Discussion

The approach is straightforward, and it can be expected that – as long as there are enough situations where the entropy for a given node (and potential pseudonym change partners) is high – this approach significantly improves the resiliency of VANET nodes to the attacks mentioned in Section V. First simulations show an improvement in the achieved level of privacy for this approach. Currently, more simulations to verify 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, for some applications, it is required that a minimum connection time is granted by the system. Routing, for example would be infeasible if node pseudonyms changed too often, because it would simply not be possible to address a node. This constraint limits the readiness for peer nodes to change pseudonyms, even if the context is good, and in turn to longer traceability of nodes. As a second consequence, it may be necessary to coordinate nodes loosely with respect to their pseudonym change.

Second, different algorithms, in particular if they take different context information into account, will output different entropy values. In order to be comparable, a reference attacker needs to be established. In addition, the more context information is considered, the fewer situations will occur where entropy is large enough to exceed the change threshold. However these situations will be better for changing. It still needs to be examined, how the impact of this better protection would be on the overall privacy of the proposed algorithm.

In a nutshell, *mix contexts* will provide an improvement of the anonymity in vehicular ad hoc networks, given a certain density of nodes.

VII. REAL WORLD ATTACKS

Up to now, the attacks consisted solely of obtaining information. After successfully carrying out the above attacks, an attacker controls information on a human and his whereabouts for a defined period. Depending on the objective and the capabilities of the attacker, this information can be (mis-)used. Here is a list of possible attacks:

- *Blackmail*. If location information contains compromising information, a third party may try to blackmail you. For

⁵The higher the entropy, the higher the anonymity

example, if you were on the premises of a competitor or have been near an establishment and someone threatens to tell your boss or wife, respectively. Or you were at another city than the one you claimed to be. Even if these incidents can be explained, they may raise suspicions, and may have serious consequences.

- *Personalized advertisements, spamming.* In the current commercial environment, companies use profiling to better address potential customers. Locations can be used to create behavioral profiles a user would not want to disclose to a particular company. *Reality Mining* [16] can be used to infer behavioral profiles, social context and others from location data. This data can be used to better address – even unwanted – advertisements to the customers.
- *Price discrimination.* As argued by Odlyzko in [17], price discrimination is probably the strongest incentive by commercial organizations to invade the privacy of people. With price discrimination, the price of an article is adjusted to the ability of an individual to pay the price or other properties. VANETs may cause pricing to be bound even to locations visited. This may imply *route discrimination*, where the price of using a certain route is calculated based on the profile of a user.
- *Suspicion by location.* There is a reported case where a firefighter has been accused of laying fire based on the profile of his at a supermarket⁶. In that particular case, the police deduced from this man buying a particular fire starter at a particular time that he must be the criminal that laid fire to a house. The fire-starter has been in jail for a couple of months after it turned out that he was in fact innocent. If full tracks of people are easily available, it is likely that this data is more frequently misused.

Note that some attacks may be limited to attackers with certain capabilities and particular knowledge. Blackmailing, for example can only be done by somebody who knows that the victim is threatened by his knowledge.

Price discrimination, on the other hand, can only be done by companies that sell goods or services; in order to price-discriminate, a name is not always necessary, a profile under a certain (permanent) pseudonym would be sufficient.

Suspicion by location can only be carried out by some authority, such as the police. In order to prevent misuse, appropriate laws and possibly also technical means are necessary.

VIII. DISCUSSION

A. Attackers

From the above analysis, it turns out that typical attackers need either appropriate legitimation, or financial resources to invade privacy in a large scale. The only cheap possibility for an individual to track a node is to use the location request mechanism in the routing protocol; for early adopter scenarios, this attack is improbable, due to the low density of the

network. However, once the network is dense enough, this attack should be taken seriously.

Another probable attack would be to install a malicious application that does the job on the target's vehicle; this requires access to the vehicle in order to install such an application. Once every vehicle is networked, this attack may become critical, similar to keylogger malware today.

Hence, in the medium term two major attackers can be identified:

- *Commercial enterprises* that seek to amend customer profiles by the aspect of tracks without the consent of a customer. They have an economic incentive to use tracks, sufficient financial resources, and existing databases to obtain enough information to carry out a successful attack.
- *Attackers within legitimate authorities* that misuse location data. These attackers have easy access to the above mentioned database of pseudonym to name mapping, and in the long run may even have access to a dense grid of receivers.

In the long term, as vehicles may more and more resemble Internet nodes, malware may become a problem, with the attacker's profile assimilating the attackers on the Internet today.

B. Attacks and Countermeasures

Table II summarizes the cost for the different attacks discussed so far. In the first phase of deployment of VANETS, the most probable attacks will be *all nodes, some place* and *some nodes, some place*. These attacks are hard to prevent and yield a list of pseudonyms for a certain location, which in turn may have some meaning that is useful for the attacker. This kind of attack can be made ineffective most easily by changing the pseudonym frequently to prevent a permanent link between a name and the pseudonym. An attacker having no legitimation to the aforementioned database would then be unable to link pseudonyms, and all aforementioned attacks would fail. Against attackers within legitimate authorities, appropriate procedures must be established that prevent the misuse of the mapping.

In summary, we propose the following measures to protect the privacy of VANET users:

- Change pseudonyms frequently,
- Change pseudonyms in appropriate contexts, called *mix-contexts*,
- Protect the use of a centralized mapping by means of laws and appropriate techniques, like a distributed mapping.

Despite these countermeasures, selected applications may need to maintain a permanent connection on top of the privacy providing communication system; we will discuss this in the next section.

C. It's not an Attack, it's a Feature

In the previous sections, the main objective was to show that a system based on permanent beaconing can be used by an attacker to invade the privacy of users. If no countermeasures

⁶http://www.schneier.com/blog/archives/2005/02/security_risks.html

Attack	Fix Cost ^a	Cost p. Month ^a	Remarks
<i>Location Tracking</i>			
All / Some nodes, some place	+	+	Hard to prevent.
Some nodes, everywhere	+	+	Feasibility and scale restricted by network.
All nodes, everywhere	+++	++++	
<i>Link Vehicle Pseudonym to User</i>			
Ask	+	+	User can decide.
Restricted Space Identification	+	+	Requires all nodes everywhere attack as a basis.
Use an existing database	+++ / ++++	+++ / ++++	Proper legitimation required
Inference from external database	+++ / ++++	+++ / ++++	Mainly cost for the external database.

TABLE II
OVERVIEW OF THE COST OF DIFFERENT ATTACKS.

^a+ – order of 100 EUR, ++ – order of 1 K EUR, +++ – order of 100 K EUR, ++++ – order of 1 Million EUR.

are taken, the system will suffer acceptance problems, and vendors integrating their system in the cars may experience decrease of their sales. On the other hand, the system will not sell without properly working applications. In fact, some of the “attacks” yield data that is useful for interesting applications; statistical data about locations and routes can be used to improve the traffic flow, the ability to resolve a location improves routing and connectivity between vehicles. Chat applications, for example, require a stable session between two cars. Finally, location data can be used to find stolen vehicles.

Profiles improve the interaction between stores and customers, and help to increase the overall efficiency of business processes; this benefits both stores and customers; for profiles to be useful, nodes must at least be recognized again.

In a nutshell, a trade-off between appropriate privacy measures and the functionality has to be found. The communication system as such, i.e. the data broadcast without protection against eavesdropping should provide the highest level of privacy possible and it must be in the hands of the user to opt for a less private system. Further, the system should support a user regaining privacy after disclosing one pseudonym. Switching off the system will not be an option, as communication will be an integral part of the safety systems of vehicles, and network operation relies on a sufficiently dense network.

IX. FUTURE WORK

Future work is to compare the attacks on changing pseudonyms in VANETs with respect to their cost. While we used financial cost in this paper for assessing attacks, that analysis will focus on algorithmic descriptions and properties. Based on this analysis, rules will be established for *mix-contexts* where vehicles can best change their pseudonyms and increase their anonymity. To compare different attacks such as the ones by Beresford and Gruteser (see Section V, a suitable metric will be derived that expresses the cost for an attacker. Further, the properties of mix contexts will be studied in simulations with focus on the trade-off between network stability and maximum traceability.

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. 01AK064.

REFERENCES

- [1] J.-P. Hubaux, S. Čapkun, and J. Luo, “The security and privacy of smart vehicles,” *IEEE Security and Privacy*, vol. 4, no. 3, pp. 49–55, 2004. [Online]. Available: <http://lcawww.epfl.ch/Publications/luo/HubauxCL04.pdf>
- [2] A. Aijaz, B. Bochow, F. Dötzer, A. Festag, M. Gerlach, R. Kroh, and T. Leinmüller, “Attacks on inter vehicle communication systems - an analysis,” The Network on Wheels Project, Tech. Rep., 2005, <http://www.network-on-wheels.de/documents.html>.
- [3] M. Gerlach, “VaneSe - An approach to VANET security,” in *Proceedings of V2VCOM 2005*, O. Altintas and W. Chen, Eds., July 2005.
- [4] F. Dötzer, “Privacy issues in vehicular ad hoc networks,” in *Workshop on Privacy Enhancing Technologies*, Cavtat, Croatia, May 2005.
- [5] B. Schneier, “Attack trees: Modeling security threats,” 1999.
- [6] P. Golle, D. Greene, and J. Staddon, “Detecting and correcting malicious data in vanets,” in *Proceedings of the first ACM workshop on Vehicular ad hoc networks*, 2004, pp. 29–37.
- [7] L. Huang, K. Sampigethaya, K. Matsuura, R. Poovendran, K. Sezaki, and M. L., “Caravan: Providing location privacy for vanet,” in *Proceedings of Escar 2005*, 2005.
- [8] M. Gruteser and D. Grunwald, “A methodological assessment of location privacy risks in wireless hotspot networks,” in *Proceedings of Security in Pervasive Computing*, 2003, pp. 10–24.
- [9] M. Gruteser and B. Hoh, “On the anonymity of periodic location samples,” in *Proceedings of Conference on Security in Pervasive Computing*, 2005.
- [10] A. R. Beresford and F. Stajano, “Location privacy in pervasive computing,” *IEEE Pervasive Computing*, pp. 46–55, 2003.
- [11] A. R. Beresford, “Location privacy in ubiquitous computing,” Dissertation, University of Cambridge, 2005.
- [12] M. Mauve, J. Widmer, and H. Hartenstein, “A survey on position-based routing in mobile ad hoc networks,” 2001. [Online]. Available: citeseer.ist.psu.edu/article/mauve01survey.html
- [13] M. Gruteser and D. Grunwald, “Anonymous usage of location based services through spatial and temporal cloaking,” in *Proceedings of the ACM MobiSys*, 2003.
- [14] A. K. Dey and G. D. Abowd, “Towards a better understanding of context and context-awareness,” in *Proceedings of 1st International Symposium on Handheld and Ubiquitous Computing*, 1999, pp. 304–307.
- [15] A. Serjantov and G. Danezis, “Towards an information theoretic metric for anonymity,” in *Lecture Notes in Computer Science*. Springer-Verlag GmbH, 2003, vol. 2482 / 2003, pp. 41 – 53.
- [16] M. I. of Technology, “Reality mining,” 2005, <http://reality.media.mit.edu/>. [Online]. Available: <http://reality.media.mit.edu/>
- [17] A. Odlyzko, “Privacy, economics, and price discrimination on the internet,” in *ICEC '03: Proceedings of the 5th international conference on Electronic commerce*. New York, NY, USA: ACM Press, 2003, pp. 355–366.
- [18] The Network on Wheels (NOW) Project, “NOW website,” 2004, <http://www.net-on-wheels.de>.