PADAVAN: Privacy-Aware Data Accumulation for Vehicular Ad-hoc Networks

Andreas Tomandl, Dominik Herrmann, Hannes Federrath
University of Hamburg
Germany
Email: lastname@informatik.uni-hamburg.de

***Abstract*—In this paper we introduce PADAVAN, a novel anonymous data collection scheme for Vehicular Ad Hoc Networks (VANETs). PADAVAN allows users to submit data anonymously to a data consumer while preventing adversaries from submitting large amounts of bogus data. PADAVAN is comprised of an n-times anonymous authentication scheme, mix cascades and various principles to protect the privacy of the submitted data itself. Furthermore, we evaluate the effectiveness of limiting an adversary to a fixed amount of messages.**

Keywords*—*VANET, Vehicular Communication, Security, Privacy, Anonymity, Data Collection

I. Introduction

In the past years vehicles evolved from purely mechanical entities to mobile computer systems with sensors that collect and process vast quantities of data. Vehicles already transmit some part of this data to third parties today. This leads to valuable comfort services like real-time traffic information, turn-by-turn directions, live fuel prices or roadside assistance. While most of the data still remains locally in the vehicle today, there is a trend to collect, process and even resell this data. In a study [1] published in January 2014 the U.S. Government Accountability Office (GAO) questioned 10 companies (six auto manufacturers, two portable navigation device companies and two navigation applications developers for mobile devices) about their in-car location-based services. All respondents stated to collect data in order to provide services for the users. Nine of them stated that they share this information with third parties to provide further services, two even for unrelated purposes (e. g., for research). One of the interviewed companies shared aggregated data with marketing firms. However, removing personally identifiable information before sharing personal data has been shown to offer less protection than expected [2], [3].

Vehicular Ad Hoc Networks (VANETs) will empower vehicles to collect and process even more vital information concerning road traffic via loose WiFi networks [4], [5]. In VANETs there are *Special Purpose Messages* (also known as Decentralized Environmental Notification Messages [6] or Roadside Alerts [7]), *Beacons* (also known as Co-operative Awareness Messages [8] or Basic Safety Messages [7]) and messages containing *Probe Data* [7]. *Special Purpose Messages* are forwarded by vehicles over multiple hops; they supply whole areas with information about specific events (e. g., an emergency vehicle approaching, a traffic jam, or an emergency braking, et cetera). Thereby, vehicles receive real-time traffic information regarding their area leading to more efficient driving. *Beacons* are transmitted frequently (e. g., five times per second) to direct neighbors (single hop only); they contain a pseudonymous vehicle identifier as well as speed, acceleration and location data. Advanced Driver Assistance Systems (ADAS) use these pieces of information to support the driver in detecting vehicles entering his blind spot, approaching with high speed or braking in front of him. Finally, vehicles may also collect sensor data (*Probe Data)* and transmit it to third parties.

The additional sensor data available in VANETs offers benefits for drivers and **data consumers** (third parties offering location-based services or scientists) alike: dangerous streets or situations can be detected based on an analysis of emergency braking data instead of accident data; streets with extensive traffic or damaged streets can be detected quickly; driver behavior before crashes might improve accident research. Moreover, VANET-specific security and privacy concepts benefit from the distributedly collected sensor data immensely (cf. Sect. IV). However, collecting this kind of data may violate the privacy of users. In this paper we propose PADAVAN *(Privacy-Aware Data Accumulation for Vehicular Ad-hoc Networks*), a framework that honors the privacy expectations of users, but still allows data consumers to obtain meaningful data. This is achieved by combining multiple building blocks:

- An n-times anonymous authentication scheme protects the privacy of the users; at the same time it secures the overall data quality for data consumers. Even if an adversary is able to transmit bogus data (e. g., via sensor manipulation), he is still limited to submitting only n data packages in a given timespan. Moreover, plausibility checks are in place to prevent the adversary from deviating considerably from typical data.
- Road Side Units (RSU) or cellular services are used to sign blinded data packages after authenticating. A RSU is able to see the pseudonym of the vehicle due to Beacon data but cannot decrypt the data packages and link it to a user even when collaborating with the data consumer. Messages are routed via mix cascades in order to prevent linkage of messages to their originator.
- In contrast to previous solutions PADAVAN enables users to control and monitor their data. Data is anonymized in the car before it is submitted to the data consumer. The use of *lean data packages* prevents adversaries from extracting information and linking data to the originator. Through the application of VANET-

specific environment information (e. g., the amount of neighbours), submitted data is kept to a minimum and the transmission frequency is dynamically adapted.

The paper is structured as follows. In Sect. II we discuss previous work, while Sect. III outlines our adversary model. The benefits of collecting data for security concepts in VANETs are presented in Sect. IV. Section V provides an overview of PADAVAN, introduces the anonymous data submission and the privacy protection mechanisms. After an evaluation of PADAVAN in Sect. VI we conclude the paper in Sect. VII.

II. Related Work

Data collection and privacy protection issues have gained broad attention in various areas of vehicular networks like VANETs and VSNs (Vehicular Sensor Networks). In VSNs, vehicles are treated as sensors and data is collected by a data consumer, which is often considered to be trustworthy. Privacy is typically provided via simple pseudonym changes. As shown in [9] this procedure is not sufficient, because vehicles submit their location too frequently (e. g., 5 times per second). Even additional privacy concepts (e. g., Mix-Zones [10], [11], Silent-Periods [12], [13] or SLOW [14]) cannot provide absolute unlinkability in the case of a global adversary as shown in [15]. With data collection services, where users provide data about themselves frequently and voluntarily in order to get access to additional services, such a global adversary is quite feasible. Therefore, further measures to protect the privacy of users, such as anonymous authentication or anonymization of data, are vital.

In [16] the authors propose to aggregate data in VANETs by using a Flajolet-Martin sketch. They carry out a simulation-based evaluation to demonstrate the efficiency of their technique. However, their analysis neglects privacy issues. The Clustered Gathering Protocol [17] for VSNs enables data consumers to aggregate data within a group of vehicles before sending it to a RSU in order to reduce the amount of traffic. Again, privacy and integrity protection is neglected in this proposal. The decentralized VESPA approach [18] enables VANET users to aggregate event data in order to learn about free parking spaces at a specific time of the day. Matrices with spatial and temporal data are shared among vehicles to improve the data; privacy issues are not considered, though. MobEyes [19], [20] enables authorities to harvest data from vehicles by using them as sensors. The authors argue that people are willing to sacrifice their privacy when their information can only be used by law enforcement authorities for purposes like tracking a suspicious vehicle via its license plates. Such measures enable the introduction of blanket surveillance, which is why they are heavily criticized by privacy advocates.

In [21] the authors propose a data collection architecture based on MACs (Message Authentication Codes). This architecture does not protect the privacy of the users against colluding infrastructure providers, though. Furthermore, the concept proposes switching keys frequently. In practice, linking these keys might still be possible due to the frequently transmitted Beacons. Moreover, the anonymous authentication that is part of this concept allows an adversary to submit an unlimited amount of bogus data to the infrastructure provider.

AnonySens [22] extends the idea of MobEyes with privacy protection techniques. Authorized applications can order mobile nodes (e. g., vehicles or mobile phones) to submit reports containing specific information. Direct Anonymous Attestation (DAA) [23] is used to authenticate anonymously and the reports contain only blurred data to maintain k-anonymity (using tessellation for location data and intervals for time data). However, AnonySens does not consider Beacon data in VANETs, i. e., an RSU might link transmitted data to a user pseudonym. Secondly, the concept enables an adversary to transmit an unlimited number of bogus reports due to the anonymous authentication. Furthermore, only location and time are obfuscated. However, other submitted data might still contain quasi-identifiers that expose the sender. Finally, the authors use historical data to calculate the tessellation for location blurring. While this might work in MANETs with slow participants, in VANETs vehicles move around fast and temporal changes in the environment (e. g., road works), might influence their movement in an area tremendously.

In [24] the authors extend AnonySens with a mix cascade. An attacker might still send unlimited amounts of bogus data and collaborating infrastructure providers might link the data to specific users.

Another approach is contained in the Dedicated Short Range Communication Standard (SAE J2735 [7]), which intends to collect *Probe Data* and submit this collected data periodically and triggered by certain events such as starts and stops. The vehicle identifier is changed every 120 s or 1 km (whatever comes first). As discussed previously, simple pseudonym changes cannot sufficiently protect the identity of users from a global adversary. Secondly, as the data is not anonymized at all, extracting information about the data source via quasi-identifiers might be feasible for adversaries.

In summary, none of the approaches mentioned above offers an adequate level of protection. The privacy of users has to be secured not only against other users but also against the data consumer that collects the submitted data. Furthermore, *pseudonymous authentication* is not sufficient in VANETs because authentication data can be linked with Beacon data. When using *anonymous authentication* it is important to keep in mind that vehicles will send out their Beacons (containing their pseudonym) frequently, even while they are authenticating. Moreover, with *anonymous authentication* the possibility to infer the identity of a user by linking context data (home and work location pairs as well as driving times) has to be taken into consideration. Therefore, data has to be anonymized and the consideration of quasi-identifiers like time and location might not be enough due to unique driving styles or the submission of whole movement patterns. Also, traffic conditions in VANETs may rapidly change and therefore pre-calculating the spatial tessellation or blurring of time periods may not suffice to maintain k-anonymity. Furthermore, constantly submitted data is difficult to anonymize without reducing its utility. Submitting less but higher quality data might be a way to get better results. Finally, to prevent adversaries from manipulating collected data, the amount of submitted data has to be restricted without losing the possibility of anonymous transmission.

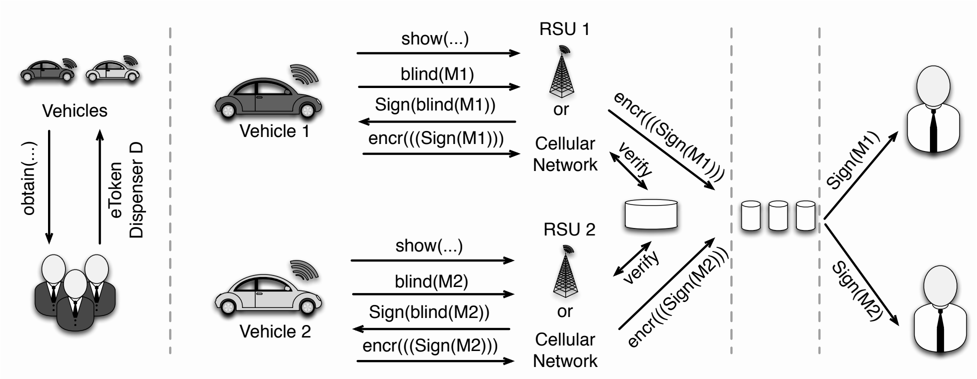

Fig. 1. Overview of PADAVAN

III. Adversary Model

Two different types of adversaries, the "data collector" and the "data forger", are considered in this paper. The "data collector" tries to maximize his knowledge about users by collecting data, creating profiles to track users and to de-anonymize data for marketing purposes or even to harm a specific user. This adversary is supposed to be able to access RSUs as well as the databases of the data consumer. Furthermore, he may apply context knowledge like data about the professional or personal lives of VANET users to link data to a specific vehicle.

The "data forger" wants to corrupt the data collected by the data consumer. He might aim to manipulate Intrusion Detection System patterns (cf. Sect. IV), re-route traffic away from his place of residence or just corrupt the data for his amusement. This adversary is supposed to be unable to access the cryptographic keys on the tamper-proof hardware of the vehicle. Instead he submits bogus data via sensor manipulation. As bogus data that deviates from typical data can be detected by outlier detection (cf. [25]), this adversary may try to submit a large amount of messages with only small deviations.

Beside the privacy threats mentioned above, data collection in VANETs also offers various benefits. In the next section we will showcase the utility of data collection for privacy and security concepts in VANETs.

IV. Benefits of Data Collection

Apart from major benefits regarding comfort services like real-time traffic routing or traffic analysis, collecting data might improve privacy and security concepts for VANETs as shown in the following two examples. First, collected data might improve the effectiveness of **privacy concepts** like mix zones. As discussed before, a simple change of pseudonyms in VANETs is not sufficient to ensure privacy. Therefore, mix zones with a specific radius are placed on junctions and vehicles will keep radio silence as well as change their pseudonyms before leaving the zone. [15] indicates that the degree of privacy might be influenced by the vehicle density and the distribution of vehicles through the different ports of a junction. To improve the placement of mix zones users can send messages stating the entry and the exit of a junction $(timestamp, x_1, y_1, x_2, y_2)$. A possible mix zone utility metric for a junction j can be obtained by calculating the entropy of all entry and exit combinations (as proposed in [10], [11]) while taking into account the relative vehicle density of the junction compared to the maximal density of all junctions in a specific area:

$$score_j = \left(-\sum_{i=1}^{N_j} p_i \cdot log_N p_i \right) + \left(\frac{V_j}{\max(V_0; V_n)} \right) \leq 2$$

Here, N represents the number of possible port combinations of the junction and p represents the probability of a vehicle choosing a specific combination; V denotes the amount of vehicles driving through a junction within a specific period of time.

Figure 2 shows simulation results of 10,000 vehicles on a map of Berlin with 10 iterations and a simulation time of 15 minutes. The experiments were conducted with VANETSim [26]. 20 randomly placed mix zones (left) as well as 20 mix zones placed according to the proposed scoring model (right) were attacked by the "advanced attack" proposed in [15]. The mix zones placed according to the scoring model provide more privacy (fraction of successful attacks between 1.6 % and 50.3 % with an average value of 11.2 %) than the randomly placed mix zones (fraction of successful attacks between 5.4 % and 80.1 % with an average value of 46.6 %). As the results show even such a simplistic scoring model can considerably increase the utility of mix zones. In further experiments the model could be improved by tweaking the relative weights of the vehicle distribution and the vehicle density. Also, other parameters like street speed limits, the influence of the time of day as well as driver behavior on specific junctions could be taken into account.

The second application we consider to demonstrate the benefit of data collected in VANETs are **intrusion detection systems**, which typically make use of explicit rules or implicit models derived by supervised learning. In order to train and evaluate such systems regular and bogus data is needed. Nearby vehicles might send received movement patterns (in form of $timestamp$, x_1, y_1, v_1, x_2, y_2, v_2, $\ldots$) from the vehicle transmitting a Special Purpose Message. Because a vehicle is supposed to react in a specific way before and after sending a Special Purpose Message (e. g., a vehicle sending an emergency braking notification is supposed to hit

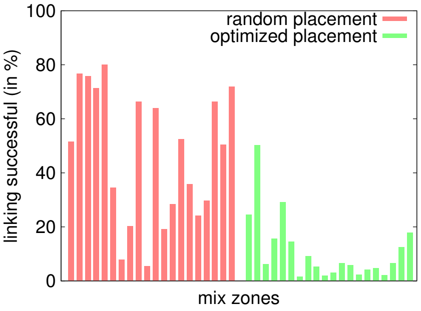

Fig. 2. Random vs. optimized mix zone placement

the brake), intrusion detection systems can be trained to detect bogus messages. Furthermore, vehicles might notice after a specific time that a message was bogus [27] and attach this information to the movement pattern. Thresholds and settings of IDS might vary depending on different cities, time of day or driver behavior. Therefore, it is vital to collect data in all areas. This might be provided through data collection in VANETs and lead to different IDS profiles for various areas.

V. PADAVAN

The following section outlines the features of PADAVAN, including its n-times anonymous authentication scheme, the anonymous communication channel and the data protection mechanisms.

A. *Overview*

Figure 1 illustrates PADAVAN. On registration vehicles are equipped with an e-token dispenser [28] as well as cryptographic keys (cf. Sect. V-B1). Vehicles authenticate themselves anonymously with an RSU (or another party through the cellular networks) in order to be able to submit n reports in a specific period of time. In order to prove the success of the authentication process to the data consumer, the vehicle presents the anonymous e-tokens to an RSU (or another third party) to obtain a blind signature [29], [30]. At this point the RSU knows the pseudonym of the authenticated vehicle (due to its Beacons) but it does not know which data was signed (cf. Sect. V-B2). The data is selected and anonymized in the vehicle to protect the privacy of the user (cf. V-C). To avoid linkage of data to a specific user in the case of collaborating RSUs or data consumers, the vehicle encrypts the signed data and submits it via a mix cascade [31] (cf. Sect V-B3).

B. *Privacy Protection in Transfer*

1) N-times Anonymous Authentication: In order to provide anonymity for users and to prevent adversaries from submitting large amounts of bogus data PADAVAN employs an anonymous authentication scheme that limits the amount of data that can be submitted by a single vehicle.

Beside other drawbacks, existing authentication schemes for VANETs like [32]–[36] do not provide a way to limit the number of anonymous submissions. This is important because, as already discussed, an attacker can manipulate sensors and send bogus data.

E-tokens and blind signatures, as used in e-cash systems [29], [30], offer n-times authentication but vehicles have to obtain new keys constantly. The scheme by Damg̊ard et al. [37] provides a more practical solution, but it has a significant impact on performance. PADAVAN employs a scheme from Camenisch et al. [28], who mandate the use of e-token dispensers that can be re-used multiple times.

In most VANET architectures each vehicle is equipped with cryptographic material during initial registration [4], [5], [38]. PADAVAN extends this existing process to set up vehicles with e-token dispensers. The issuer retains tracing information as well as revocation information (as shown in Fig. 1 Step 1). The e-token dispenser can be used by a vehicle to authenticate to a verifier n times in a specific time period. The tracing information is used to calculate token serial numbers for each user in order to prevent users from continuously flooding the data consumer with replayed messages. The revocation information is used to revoke specific e-token dispensers in case of detected misbehavior. Note that tracing users or revoking their dispensers is only possible if users authenticate with more than n e-tokens within the designated time period.

To submit data a vehicle will spend an e-token to authenticate itself to an RSU (or to a service provided via the cellular network) as shown in Step 2 in Fig. 1. As a result the vehicle receives an updated e-token dispenser while the verifier obtains a transcript and an e-token serial number, which is stored in a database to allow other verifiers to check for double spending. Plausibility checks comparing the amount of e-token serial numbers with the amount of signed messages by a verifier prevent malicious verifiers from the transmission of own messages with their key. Using two different transcripts from submissions of the same e-token serial number, verifiers can prove to the issuer that a vehicle tried to double-spend an e-token. In this case tracing as well as revocation will be possible.

This authentication process enables vehicles to anonymously authenticate to a verifier, but on its own this feature is not sufficient: As discussed in Sect. II vehicles constantly broadcast Beacons that contain their pseudonymous vehicle ID. In the following two sections we will describe how to prevent adversaries from linking information from Beacons with the data submitted to the data consumer.

2) Anonymous Proof of Authentication: To guarantee the privacy of users it is necessary to prove to the data consumer that a user has authenticated successfully without revealing both, his identity and the submitted data. After authenticating the verifier could sign the data but a vehicle would have to encrypt the data with the public key of the data consumer to hide it from the verifier. This would lead to a privacy breach if the RSU and the data consumer collaborated. One way to sign data without revealing the content to a verifier is the blind signature scheme [29], [30]. After authenticating as described in Sect. V-B1 a vehicle blinds the data and the verifier signs it with its private key. Since the data already contains obfuscated location data, the fact that a specific key can be linked to a specific area poses no additional privacy threat. The vehicle unblinds the data in order to obtain the signed message of the verifier (as shown in step 2 in Fig. 1). In PADAVAN each vehicle is allowed to sign m data sets after authenticating to a verifier one time, to improve the performance and to reduce the number of cryptographic operations. The e-token for every data set is signed independently to prevent linkability and a fixed amount of data in one package is defined to prevent adversaries to circumvent the restriction. A data package with more data than permitted is rejected by the data consumer.

However, the signed data could still be linked to the vehicle, if submitted directly and instantly to the data consumer. This issue is resolved by the anonymous data transfer scheme presented in the following.

3) Anonymous Data Transfer: In order to provide anonymous data transfer a mix cascade [31] is used to submit data to the data consumer (as shown in step 3 in Fig. 1). As long as at least one mix node is not collaborating with the verifier and the data consumer, an adversary cannot link the data to the vehicle. After unblinding the signed message from the verifier a vehicle will lookup a mix route and encrypt the message. Besides the signed message a data packet might contain an optional anonymous return address with the pseudonym of the vehicle. This address might be added to technical support requests or when fully autonomous vehicles are to be monitored.

C. Privacy Protection of Data

To protect the privacy of a user, anonymous data transfer is still not sufficient. Even anonymized data without identifiers might be linked to a person: For instance, Sweeney [39] showed that approximately 87% of the population of the United States can be uniquely identified solely by the attributes “gender”, “date of birth” and the “zip code”. In the following we describe the principles to protect submitted data in PADAVAN.

PADAVAN implements three principles to protect the submitted data: “Data Splitting”, “Data Retention” and “User Control”

A common approach to accumulate data in VANETs consists in vehicles collecting data for a specific time and submitting them as a data package to the data consumer, who will derive various pieces of information from the data. The “Dedicated Short Range Communication” (DSRC) Implementation Guide [7] defines this data as *Probe Data*. This approach requires a high level of trust in the data consumer, even when mechanisms for anonymous data transfer are in place. Given a large contiguous data set the data consumer may be able to identify its originator. Therefore, PADAVAN implements the principle of **“Data Splitting”**: Vehicles transmit singular packages for each use case (e. g., timestamp, location and speed for real-time traffic services; emergency braking information and location data for dangerous streets analysis).

The second principle of PADAVAN is **“Data Retention”**. Instead of transmitting data frequently, in PADAVAN vehicles analyze the data and adapt the rate at which submissions occur. Due to Beacons, the vehicle knows the amount of neighbors and attaches the amount to data packages. A message regarding real-time traffic information containing timestamp t, location x, y, speed v and amount of neighbors n carries the same information as n messages with t, x, y, v. Furthermore, vehicles vary the time of submission probabilistically to prevent inference attacks.

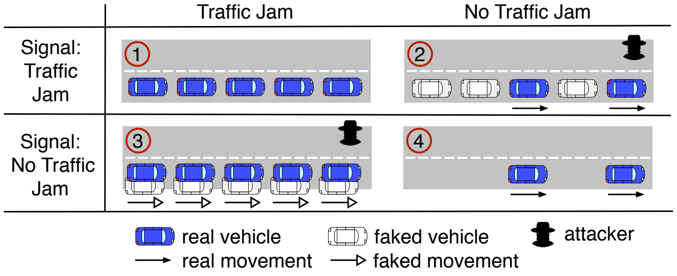

Fig. 3. Traffic jam attack

The last principle, **“User Control”**, consists in user-side anonymization, selection and overview of data. Other concepts (e. g., [21], [22]) rely on the VANET infrastructure or third parties to anonymize the submitted data. This protects against external adversaries, but not against (collaborating) insiders. In PADAVAN vehicles apply obfuscation and anonymization techniques (e. g., [40], [41]) before sending the data to the data consumer. Before data consumers may share collected data with other third parties, they have to apply further anonymization techniques [22]. Besides the anonymization of data, users are able to inspect the collected data or to opt out of collection and submission altogether. In some special cases vehicles may be required to send data about other vehicles (e. g., information for intrusion detection systems, cf. Sect. IV, or due to legal monitoring requirements regarding autonomous vehicles). To indicate to other vehicles the desired level of privacy, a vehicle could attach a “do not track” flag to its Beacons.

VI. Evaluation

In the following we evaluate the effectiveness of the n-times authentication scheme to limit the influence of bogus data. In a real-world deployment various parameters must be chosen to guarantee optimal protection against adversaries, while still maintaining practicability. For instance, the value of n (the number of times a vehicle can authenticate anonymously) depends on the amount of services that are operated within PADAVAN. Therefore, we cannot provide universally suitable values for all parameters in this paper. Instead, we will focus on two concrete case studies, a real-time traffic service and the placement of mix zones in order to demonstrate that the impact of adversaries is limited in PADAVAN regardless of the specific value of n.

Figure 3 illustrates two extreme situations in real-time traffic, a traffic jam and a clear road. An adversary might try to convince the data consumer of the opposite event in each case. To do so, he must provide more real-time traffic data than the benign vehicles. The ratio between bogus and benign data is governed by the following formula:

$$s = \frac{n \cdot m \cdot A}{V \cdot t \cdot \frac{n \cdot m}{T_{\text{Slot}}}} = \frac{T_{\text{Slot}} \cdot A}{V \cdot t}$$

The strength of an attacker is given by s; for $s > 1$ the adversary has submitted more data than the benign vehicles. V represents the vehicle amount per minute that passes through a street segment (1 km with an average speed of 60 $\frac{km}{h}$ in this example). The amount of messages a vehicle can send per time slot T_{Slot} is given by n, the amount of possible authentications, and m, the number of packages a verifier signs given a single authentication token. A indicates the amount of collaborating adversaries. The last factor t is the time vehicles must stand still until a traffic jam is detected by the data consumer.

As a first important finding, this formula indicates that the permissible number of anonymous messages within a time slot, i. e., the values of n and m, does not influence the strength of the adversary. Only relevant are V, t, T_{Slot} and A. V, the vehicle density cannot be influenced. In Scenario 2 of Fig. 3 the density of real vehicles might be quite small on some streets (e. g., $t = 10$, $T_{\text{Slot}} = 60$, $A = 1$ requires at least 6 vehicles per minute to be save). In the third scenario, when an adversary tries to dissolve a traffic jam, the street is filled with benign vehicles and V is very high (e. g., $t = 10$, $T_{\text{Slot}} = 60$, $V = 100$, requires at least 4 attackers collaborating). The factor t can be influenced by the data consumer and might be increased in areas with low vehicle densities. To weaken the attacker further T_{Slot} might be chosen as low as possible. Further plausibility checks by the data consumer like "did a high vehicle density suddenly appear", can weaken an attacker even more.

As a second example, an attack on the optimized mix zone placement introduced in Sect. IV will be discussed. One objective of an attacker might be to improve the score for a junction with a bad score. This is almost impossible because a junction with a bad score will typically have a low vehicle density and therefore the attacker will have to submit a large amount of bogus data (which is infeasible due to the n-times authentication scheme). Another objective, i. e., to decrease the score of a well-rated junction to be able to track vehicles in this area, requires less data and is easier for an attacker. In the following, the second objective with the stronger adversary will be analyzed.

Benign vehicles M_{BV} and attacker vehicles M_{AV} generate messages according to the following formulas:

$$M_{BV} = V \cdot T_{\text{Slot}} \cdot T_{\text{Experiment}} \cdot \frac{n \cdot m}{J}$$

$$M_{AV} = n \cdot m \cdot T_{\text{Experiment}} \cdot A$$

$T_{\text{Experiment}}$ represents the simulation time of an experiment while J is the amount of junctions a vehicles will pass on average in T_{Slot}.

As in the real-time traffic scenario (cf. Fig. 3), the number of messages per time slot ($n \cdot m$) has no impact on the strength of the attacker. The most important parameter is simply A, the number of (collaborating) attacking vehicles. Figure 5 shows the attack on an existing mix zone (with an entropy of ≈0.48 and a relative vehicle density of ≈0.36; selected from the analyzed junctions of Berlin in Sect. IV; score ranking 1047 of 1163; $n \cdot m = 20$; $V = 20$, $J = 100$, $T_{\text{Experiment}} = 24$). At first, an attacker trying to decrease the score will in fact *increase* it. This is the case because the attacker increases the relative density of the mix zone by submitting additional messages. The decrease of the entropy is not sufficient to decrease the overall score. When the density reaches its maximum of 1 the total score begins to decrease, but does not fall below the baseline. Even if the density would be 1 at the outset, 50 collaborating attackers have only a limited influence. As the distribution of all scores in the map extract of Berlin (cf. Fig. 4) shows, even a mix zone score of 1 in total would be one of the higher scores. Therefore, the influence of an attacker to decrease the overall score of a mix zone is quite limited.

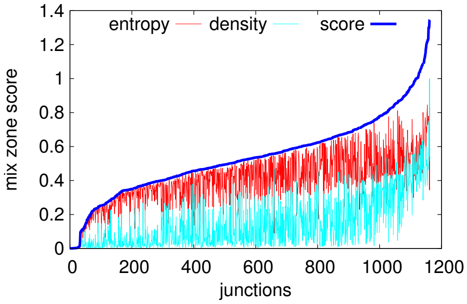

Fig. 4. Mix zone score distribution

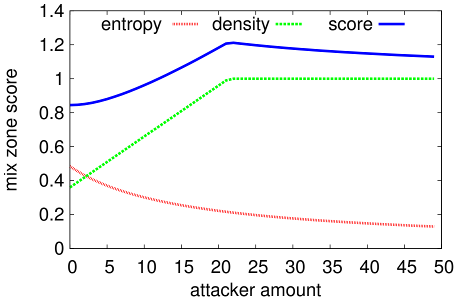

Fig. 5. Mix zone attack

VII. Conclusion

In this paper we present PADAVAN, a novel data collection framework for vehicular networks. PADAVAN is optimized to meet the privacy expectations of car drivers, while at the same time maintaining a high degree of data quality for data consumers.

To the best of our knowledge PADAVAN is the first data collection framework to employ n-times anonymous authentication. This building block limits the amount of bogus data an adversary can introduce into the system. The use of blind signatures combined with a mix-cascade ensures the privacy of users, even if infrastructure providers collaborate. PADAVAN offers complete user control including the possibility to opt out; users are not required to trust third parties as all data obfuscation and anonymization happens within the vehicle. Through the application of vehicle density information, collected data is kept to a minimum and small-sized anonymous data packages provide a high level of unlinkability.

In future work we plan to further analyze the utility of the collected data for the training of intrusion detection profiles as well as for the optimal placement of mix zones.

Acknowledgment

The authors would like to thank Christoph Gerber for his helpful comments.

References

[1] “IN-CAR LOCATION-BASED SERVICES: Companies Are Taking Steps to Protect Privacy, but Some Risks May Not Be Clear to Consumers. GAO-14-81: Published: Dec 6, 2013. Publicly Released: Jan 6, 2014.”

[2] M. Barbaro, T. Zeller, and S. Hansell, “A face is exposed for AOL searcher no. 4417749,” *New York Times*, vol. 9, no. 2008, 2006.

[3] A. Narayanan and V. Shmatikov, “How to Break Anonymity of the Netflix Prize Dataset,” *arXiv preprint cs/0610105*, 2006.

[4] M. Raya and J. Hubaux, “The Security of Vehicular Ad Hoc Networks,” in *Proceedings of the 3rd ACM workshop on Security of ad hoc and sensor networks (SASN)*. ACM, November 2005, pp. 11–21.

[5] K. Plößl and H. Federrath, “A Privacy Aware and Efficient Security Infrastructure for Vehicular Ad Hoc Networks,” in *Security in Information Systems: Proceedings of the 5th International Workshop on Security in Information Systems – WOSIS 2007*, 2007.

[6] “ETSI TS 102 637-3: Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Application; Part 3: Specification of Decentralized Environmental Notification Basic Service.” 2010.

[7] “SAE J 2735. Dedicated Short Range Communications (DSRC) Message Set Dictionary,” 2009.

[8] “ETSI TS 102 637-2: Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Part 2: Specification of Co-operative Awareness Basic Service.” 2011.

[9] B. Wiedersheim, Z. Ma, F. Kargl, and P. Papadimitratos, “Privacy in Inter-Vehicular Networks: Why simple pseudonym change is not enough,” in *2010 Seventh International Conference on Wireless On-demand Network Systems and Services (WONS)*. Los Alamitos: IEEE Computer Society Press, Februar 2010, pp. 176–183.

[10] A. R. Beresford and F. Stajano, “Location Privacy in Pervasive Computing,” *IEEE Pervasive Computing*, vol. 2, 2003.

[11] ——, “Mix Zones: User Privacy in Location-aware Services,” in *2nd IEEE Conference on Pervasive Computing and Communications Workshops (PerCom 2004 Workshops), 14-17 March 2004, Orlando, FL, USA*. IEEE Computer Society, 2004, pp. 127–131.

[12] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, “Enhancing wireless location privacy using silent period,” in *Proceedings of the 2005 IEEE Wireless Communications and Networking Conference (WCNC)*, New Orleans, LA, USA, March 2005, pp. 1187–1192.

[13] L. Huang, H. Yamane, K. Matsuura, and K. Sezaki, “Silent Cascade: Enhancing Location Privacy Without Communication QoS Degradation,” in *Security in Pervasive Computing, Third International Conference, SPC 2006, York, UK, April 18-21, 2006, Proceedings*, ser. Lecture Notes in Computer Science, vol. 3934. Springer, 2006, pp. 165–180.

[14] L. Buttyán, T. Holczer, A. Weimerskirch, and W. Whyte, “SLOW: A Practical Pseudonym Changing Scheme for Location Privacy in VANETs,” in *Proceedings of the IEEE Vehicular Networking Conference (VNC), Tokyo, Japan, 2009*. IEEE, October 2009.

[15] A. Tomandl, F. Scheuer, and H. Federrath, “Simulation-based evaluation of techniques for privacy protection in VANETs,” *Wireless and Mobile Computing, Networking and Communications (WiMob)*, 2012.

[16] C. Lochert, B. Scheuermann, and M. Mauve, “Probabilistic aggregation for data dissemination in VANETs,” in *Proceedings of the 4th ACM international workshop on vehicular ad hoc networks*, 2007.

[17] I. Salhi, M. O. Cherif, and S.-M. Senouci, “A new architecture for data collection in vehicular networks,” in *Communications, 2009. ICC’09. IEEE International Conference on*, 2009.

[18] B. Defude, T. Delot, S. Ilarri, J.-L. Zechinelli, and N. Cenerario, “Data aggregation in VANETs: the VESPA approach,” in *Proceedings of the 5th Annual International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*, 2008.

[19] U. Lee, B. Zhou, M. Gerla, E. Magistretti, P. Bellavista, and A. Corradi, “Mobeyes: smart mobs for urban monitoring with a vehicular sensor network,” *Wireless Communications, IEEE*, 2006.

[20] U. Lee, E. Magistretti, M. Gerla, P. Bellavista, and A. Corradi, “Dissemination and harvesting of urban data using vehicular sensing platforms,” *Vehicular Technology, IEEE Transactions on*, 2009.

[21] B. Hoh, M. Gruteser, H. Xiong, and A. Alrabady, “Enhancing security and privacy in traffic-monitoring systems,” *Pervasive Computing, IEEE*, vol. 5, no. 4, pp. 38–46, 2006.

[22] A. Kapadia, N. Triandopoulos, C. Cornelius, D. Peebles, and D. Kotz, “AnonySense: Opportunistic and privacy-preserving context collection,” in *6th International Conference on Pervasive Computing*. Springer, 2008.

[23] E. Brickell, J. Camenisch, and L. Chen, “Direct anonymous attestation,” in *Proceedings of the 11th ACM Conference on Computer and Communications Security*, 2004.

[24] K. Huang, S. Kanhere, and W. Hu, “Preserving privacy in participatory sensing systems,” *Computer Communications*, vol. 33, no. 11, pp. 1266–1280, 2010.

[25] V. Hodge and J. Austin, “A survey of outlier detection methodologies,” *Artificial Intelligence Review*, vol. 22, no. 2, pp. 85–126, 2004.

[26] A. Tomandl, D. Herrmann, K.-P. Fuchs, H. Federrath, and F. Scheuer, “VANETsim: An open source simulator for security and privacy concepts in VANETs,” *to be published at 9th International Workshop on Security and High Performance Computing Systems (SHPCS 2014)*, 2014.

[27] A. Tomandl, K.-P. Fuchs, and H. Federrath, “REST-Net: A rule-based IDS for VANETs,” *to be published at Wireless and Mobile Networking Conference (WMNC 2014)*, 2014.

[28] J. Camenisch, S. Hohenberger, M. Kohlweiss, A. Lysyanskaya, and M. Meyerovich, “How to win the clonewars: efficient periodic n-times anonymous authentication,” in *Proceedings of the 13th ACM conference on Computer and Communications Security*, 2006.

[29] D. Chaum, “Blind signatures for untraceable payments,” in *Advances in cryptology*. Springer, 1983.

[30] ——, “Blind signature system,” in *Advances in cryptology*. Springer, 1984, pp. 153–153.

[31] ——, “Untraceable electronic mail, return addresses, and digital pseudonyms,” *Communications of the ACM*, vol. 24, no. 2, pp. 84–90, 1981.

[32] R. Lu, X. Lin, H. Zhu, P. Ho, and X. Shen, “ECPP: Efficient conditional privacy preservation protocol for secure vehicular communications,” in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, 2008.

[33] X. Lin, X. Sun, X. Wang, C. Zhang, P. Ho, and X. Shen, “TSVC: timed efficient and secure vehicular communications with privacy preserving,” *Wireless Communications, IEEE Transactions on*, 2008.

[34] A. Perrig, R. Canetti, J. Tygar, and D. Song, “The TESLA broadcast authentication protocol,” *CryptoBytes*, vol. 5, no. 2, pp. 2–13, 2002.

[35] V. Paruchuri and A. Durresi, “PAAVE: Protocol for Anonymous Authentication in Vehicular Networks using Smart Cards,” in *Global Telecommunications Conference (GLOBECOM 2010), 2010 IEEE*, 2010.

[36] A. Studer, E. Shi, F. Bai, and A. Perrig, “TACKing together efficient authentication, revocation, and privacy in VANETs,” in *Sensor, Mesh and Ad Hoc Communications and Networks, 2009. SECON’09. 6th Annual IEEE Communications Society Conference on*. IEEE, 2009, pp. 1–9.

[37] I. Damgård, K. Dupont, and M. Østergaard Pedersen, “Unclonable group identification,” in *Advances in Cryptology-EUROCRYPT 2006*. Springer, 2006, pp. 555–572.

[38] G. Calandriello, P. Papadimitratos, J. Hubaux, and A. Lioy, “Efficient and Robust Pseudonymous Authentication in VANET,” in *Proceedings of the Fourth International Workshop on Vehicular Ad Hoc Networks, VANET 2007, Montréal, Québec, Canada, September 10, 2007*, W. Holfelder, P. Santi, Y.-C. Hu, and J.-P. Hubaux, Eds. ACM, 2007, pp. 19–28.

[39] L. Sweeney, “Uniqueness of Simple Demographics in the U.S. Population,” Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, PA, Tech. Rep. LIDAP-WP4, 2000.

[40] R. Shokri, J. Freudiger, and J. Hubaux, “A unified framework for location privacy,” *3rd Hot Topics in Privacy Enhancing Technologies (HotPETs)*, pp. 203–214, 2010.

[41] S. Gambs, M. Killijian, and M. del Prado Cortez, “Show me how you move and i will tell you who you are,” in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Security and Privacy in GIS and LBS*. ACM, 2010, pp. 34–41.