Nowhere to hide? Mix-Zones for Private Pseudonym Change using Chaff Vehicles

Christian Vaas*, Mohammad Khodaei†, Panos Papadimitratos†‡, Ivan Martinovic*
*System Security Lab, University of Oxford, United Kingdom
{christian.vaas, ivan.martinovic}@cs.ox.ac.uk
†Networked Systems Security Group, KTH Royal Institute of Technology
{khodaei, papadim}@kth.se
‡RISE SICS, Stockholm, Sweden

***Abstract*—In vehicular communication systems, cooperative awareness messages provide contextual information required for transportation safety and efficiency applications. However, without the appropriate design, these messages introduce a new attack vector to compromise passenger privacy. The use of ephemeral credentials – *pseudonyms* – was therefore proposed, essentially to split a journey into unlinkable segments. To protect segment transitions, encrypted mix-zones provide regions where vehicles can covertly change their pseudonyms. While previous work focused on the placement, shape, and protocols for mix-zones, attacks that correlate vehicles entering and existing these zones still remain a problem. Furthermore, existing schemes have only considered homogeneous traffic, disregarding variations in vehicle density due to differences in driver population, road layout, and time of day. Without realistic experimental results, any conclusion on real-world applicability is precarious. In this paper, we address this challenge and present a novel scheme that works independent of vehicles' mobility patterns. More precisely, our system generates fictive *chaff* vehicles when needed and broadcasts their traces, while it remains unobtrusive if sufficiently many vehicles are present. This greatly improves privacy protection in situations with inherently low traffic density, e.g., suburban areas, and during low traffic periods. Our scheme ensure that an external attacker cannot distinguish between real and chaff vehicles, while legitimate vehicles can recognize chaff messages; this is important, because chaff vehicles (and messages) must not affect the operation of safety applications. In our evaluation, we compare our chaff-based approach with an existing cryptographic mix-zone scheme. Our results under realistic traffic conditions show that by introducing fictive vehicles, traffic flow variations can be smoothed and privacy protection can be enhanced up to 76%.**

I. Introduction

Vehicular Ad-hoc Networks (VANETs) are to play a key role in the development of intelligent transportation systems. While recent technological advances, such as LIDAR increased the sensing capabilities of individual vehicles, methods such as cooperate collision avoidance and adaptive cruise control leverage collectively gathered information disseminated throughout a VANET. To achieve this, vehicles' On-Board Units (OBUs) periodically broadcast Cooperative Awareness Messages (CAMs) that provide information about their exact location, heading, and velocity in real time. Digital signatures using public key cryptography ensure the authenticity, integrity, and non-repudiation of messages. According to international standards (IEEE 1609.2 WG [1] and ETSI [2]), each OBU, authorized to participate in the vehicular communication system, is equipped with public-private keys pairs. Using long-term credentials for signature generation allows an attacker to track a victim throughout its journey. As a solution for privacy preservation, ephemeral credentials called *pseudonyms* were proposed [3], [4]. For credential provisioning, schemes like SECMACE [5], [6] provide a Vehicular Public-Key Infrastructure (VPKI) consisting of several Certificate Authorities (CAs). By frequently changing to a new pseudonym, these schemes aim to break a journey down into smaller, unlinkable segments.

At the same time, it has been understood that messages signed under different pseudonyms can be linked by an eavesdropper; either by examining the pseudonyms themselves or by basing inferences on the content of the signed messages [7]. *Syntactic* and *semantic* linking [7], [8] exploit the circumstances under which vehicles change pseudonyms. Based on the time of a transition, an attacker might especially observe an isolated pseudonym change, and associate the old and new identifiers through *syntactic linking* [8], [9]. During a *semantic linking* attack, the adversary uses physical constraints of the road layout, velocity, and heading of a victim's vehicle to predict its trajectory and link pseudonyms [7], [10]. While appropriate pseudonym provisioning policies alleviate syntactic linking [5], [6], [11], privacy protection is lowered by semantic linking[1] [7], [8], [10], [13], [14].

Different solutions have been proposed to thwart these linking attacks. In silent period-based schemes, vehicles refrain from message transmission while changing pseudonyms [8], [15], [16]. However, the resulting lack of situational awareness diminishes the reliability of safety applications, such as collision avoidance [17], and greatly increases the probability of an accident [18]. Alternatively, Cryptographic mix-zones (CMIXs) cover road segments in which messages are encrypted to conceal pseudonym changes and CAMs [19], [20]. Improved privacy protection was shown under idealized conditions, i.e., homogeneous traffic and simplified road layouts. However, the resiliency of CMIX-based schemes against link-

[1]Note that connecting such anonymous location profiles to real identities of vehicle owners is the final step, e.g., tracing their commutes and identify home/work locations [12], or full de-anonymization of vehicles by *honest-but-curious* VPKI entities [5], [6], [9].

ing attacks highly depends on the driver population, vehicle arrival rates, and a uniform transition probability for vehicles traversing the mix-zones. More precisely, under realistic mobility patterns [21], real-world road network topologies [7], and mix-zone geometries [22], an attacker could infer enough information to harm user privacy [7], [19], [23]. Hence, these patterns have a significant impact on the efficiency and resiliency of such schemes in real-world scenarios [24].

In this paper, we address these problems by providing a novel mix-zone scheme with protection against syntactic and semantic linking attacks. More precisely,

- We propose a scheme that introduces fictive *chaff* vehicles without interfering with the operation of existing safety applications. Our scheme smooths variations in road layout, time of day, and vehicle mobility patterns to improve user privacy. Unlike previous work, our construction considers the impact of a network of mix-zones on safety applications instead of one isolated mix-zone.
- We provide missing experimental results to show the impact of linking attacks by an eavesdropping adversary who uses features of the traffic flow. This completes the study by [14], which did not consider such attacks.
- In a quantitative analysis, we compare our chaff-based approach with an existing CMIX scheme to show the impact of variations in traffic density on passenger privacy. Our results under realistic traffic conditions show that by introducing fictive vehicles, we can enhance user privacy protection up to 76%.

In the rest of the paper, we survey the state-of-the-art research efforts (Sec. II) and describe the system model, assumptions and adversarial model (Sec. III). We present our chaff-based CMIX scheme (Sec. IV), followed by a qualitative and quantitative analysis of security and privacy (Sec. V and VI). We then evaluate the performance of our scheme (Sec. VII) and conclude the paper (Sec. VIII).

II. Related Work

Due to the openness of wireless transmissions, an observer can arbitrarily eavesdrop on VANET communication [25]–[27]. With advances in broadcast technology to extend the transmission range of OBUs [28], VANET messages become increasingly accessible for an attacker. This information allows semantic linking attacks which rely on location and heading information of continuously broadcast CAMs [7].

Different pseudonym transition strategies to prevent an attacker from inferring such information have been proposed. To evade correlation attacks, some proposals suggest that each vehicle should turn its wireless transmitter off for a randomly chosen interval and change pseudonym within that silent period [16], [29], [30]. Even though such schemes could improve user privacy, they impose a performance penalty on safety applications [18]. As a mitigation, vehicles could change their pseudonyms when their speed drops below 30 km/h [8]. But, an adversary can still conduct syntactic linking due to a lack of synchronization among vehicles [9], or track vehicles across pseudonym changes by predicting their trajectories [15].

Another line of study that does not impair safety applications proposes encrypted pseudonym transitions. In CMIX-based schemes [19], a cryptographic mix-zone is created at appropriate times and places, e.g., at intersections. When crossing these regions, vehicles change their pseudonyms privately while their communication is encrypted, which prevents syntactic and semantic linking attacks. However, the achieved privacy protection highly depends on the number of vehicles participating in the mix-zone. Consequently, user privacy is degraded under low traffic density, e.g., in a highway scenario [31]. Moreover, an attacker could compromise pseudonym changes within a mix-zone based on the traffic mobility pattern and vehicle speed [32]. Unlike such schemes, we provide privacy protection regardless of variations in road layout, time of day, vehicle density, and mobility patterns.

An alternative to these static mix-zones, dynamic formation of mix-zones were proposed [33]. Each OBU is provided with a global symmetric key, which it uses to spontaneously initiate an encrypted area for pseudonym change. To prevent semantic linking, such an approach requires vehicles to change their speed, lane, or direction when updating their pseudonyms, bringing the practicality of this scheme into question. Since the scheme allows any OBU to terminate encryption periods, an internal attacker could use this to reduce the anonymity set, thus, compromising user privacy. Additionally, the scheme introduces significant overhead on key management: for each revocation event, a new symmetric key has to be distributed to all legitimate vehicles within the system.

III. System Model and Objectives

A. System Model and Assumptions

We assume a VPKI with Long Term CA (LTCA) and Pseudonym CA (PCA) [6], [34], [35]. The LTCA provides registered vehicles with a Long Term Certificate (LTC), used to authorize the acquisition of pseudonyms from a PCA [5], [6], [11]. Road-Side Units (RSUs) are authorized to request a set of chaff pseudonyms equivalent to the pseudonyms used by OBUs. The state-of-the-art VPKI deployments can easily be adapted to provide pseudonyms in a timely manner [35]. All VANET entities are loosely synchronized with the VPKI clock. We further assume that all OBUs and RSUs are in possession of a Hardware Security Module (HSM). Private keys stored in an HSM cannot be extracted and only one pseudonym can be active at a time for OBUs.

Messages are signed using the OBU's private key corresponding to the currently valid pseudonym. In case a faulty behavior is detected, authenticated messages can be used by a Resolution Authority (RA) to retrieve the long-term identity of the vehicle [6]. The misbehaving OBU is subsequently excluded from further participation in the system by invalidating its credentials, added to a revocation list [36].

All RSUs are provided with a wired connection to communicate with VPKI entities and other RSUs. These RSUs are aware of their locality, the road layout, and their neighboring RSUs. RSUs relay requests between OBUs and VPKI entities. Unless revoked, OBUs and RSUs are considered trusted.

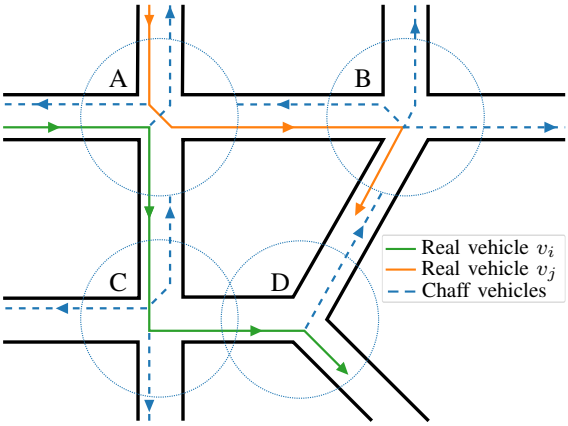


Fig. 1: Network of four mix-zones: A, B, C, and D. Their encrypted regions are indicated with dotted blue lines. Green and orange solid lines show two real vehicles; each dashed blue line denotes the trace of a chaff vehicle.

B. Adversary Model

We adhere to the adversarial model defined in the literature [3], [4]: internal adversaries, i.e., faulty, compromised, or malicious, or external ones, i.e., unauthorized entities. Internal adversaries have credentials and cryptographic material making them a legitimate part of the system; they can deviate from system protocols. External adversaries do not possess valid credentials; but they can affect the system operation. From a different perspective, adversaries could be active, i.e., capable of injecting or modifying exchanged messages, or passive, i.e., capable of collecting information by eavesdropping the communication. In this work, we consider passive external adversaries with wireless receivers placed along a road network to eavesdrop on vehicular communications. We assume an attacker who deploys off-the-shelf receivers. Performing OBU fingerprinting or angle-of-arrival computation requires significantly more sophisticated equipment and it is beyond the scope of this investigation.

The attacker passively eavesdrops the communication and acquires information to compromise passenger privacy. Such information can be derived from CAMs, e.g., timing, velocity, heading, and location. Attackers could also collude and share individually collected information to compromise user privacy. Leveraging this information, they try to perform semantic linking by associating old and new pseudonym after a pseudonym transition. Especially because our scheme must not impair safety applications, a mechanism is provided for OBUs to distinguish between fictive and real vehicle messages. This poses an attractive target for an attacker, given knowledge of our scheme, who might try to filter chaff messages out.

IV. Chaff-based CMIX

A. System Overview

To provide a solution resilient to variations in traffic density, we introduce fictive *chaff vehicles*. While these vehicles do not exist in reality, *chaff CAMs* are generated and signed with the private keys of *chaff pseudonyms* to materialize chaff vehicles' presence. The purpose is to decrease the chances of an attacker compromising a victim's pseudonym change by linking its old to its new pseudo-identity. Our scheme uses designated RSUs, positioned at locations where vehicles physically mix, e.g., at intersections [19], to create encrypted mix-zones for private pseudonym change. RSUs serve three purposes: maintaining a symmetric key to establish the encrypted region, generating messages resembling chaff vehicles, and ensuring that chaff pseudonyms do not interfere with safety applications.

Using the scenario in Fig. 1, we describe different tasks of our scheme. The four mix-zones (A, B, C, and D) indicated with dotted circles denote the encrypted regions of the road network. For each mix-zone, an RSU manages chaff vehicle generation and key distribution for vehicles in proximity.

After a vehicle entered a mix-zone, an RSU uses its knowledge about other vehicles in that zone, their intended trajectory, and the road layout to determine how many chaff vehicles are required. In the case of mix-zone A, two vehicles, indicated in orange and green, enter at the same time. Given the two real vehicles, their actual trajectories, and mix-zone exit points, two chaff vehicles need to be created to fill the remaining two exits (blue dashed lines); in fact, pretending that two 'real' vehicles exit from those exit points. The RSU responsible for mix-zone A uses location and heading from messages of previously observed real vehicles to generate synthetic CAMs, resembling the traces towards adjacent mix-zones. Message timestamps are adapted to emulate authentic location information and velocity, making chaff and real vehicles indistinguishable. While RSUs broadcast CAMs within their range directly, broadcasting chaff messages outside their coverage needs to be delegated[2]. Neighboring RSUs can undertake messages within their coverage while messages that are supposed to originate outside any RSU coverage can be delegated to real vehicles.

In order to preserve the correct functionality of safety applications, our scheme provides vehicles with information to identify chaff messages. Therefore, each RSU keeps a Cuckoo Filter (CF) [37] which it distributes to OBUs. This data structure includes fingerprints of chaff pseudonyms used to sign chaff CAMs from adjacent mix-zones. When receiving a CAM, an OBU tests the attached pseudonym's fingerprint against its CF; if positive, the message is discarded. In our example scenario, the RSU, operating mix-zone B, maintains a CF that includes chaff pseudonyms used by the zones A, D and B itself. Similarly to Bloom Filters, CFs provide fast membership tests at the cost of a false positive rate, but in contrast support dynamic updates of the underlying set. This also alleviates the introduced computational overhead.

Yet, before a pseudonym can be used, neighboring RSUs and OBUs in transit between mix-zones have to be notified to update their CF. Due to propagation delays, these updates may inflict a delay until a chaff pseudonym can be used. However, to ensure effective privacy protection, an RSU cannot delay chaff message emission. To mitigate these propagation delays,

[2]We note that the encryption radius of a mix-zone is not equivalent to an RSU's transmission range. The effect of these radii and their impact on privacy is evaluated in Section VII.

TABLE I: Protocol parameters summary and description.

C_{p^+}	Certificate corresponding to public key p^+
CF_e	Entity e's local Cuckoo Filter CF
$E_k(msg)$	Encryption of message msg using key k
$FP(\cdot)$	Fingerprint of a chaff pseudonym
LK_e^+, LK_e^-	Long-term public and private key of entity e
m	Number of vehicles in transit between two mix-zones
PK_e^+, PK_e^-	Pseudonymous public and private key pair of entity e
Pos_e	Position information (longitude, latitude)
R_r^{MIX}	Encryption radius of mix-zone r
Sig_{k^-}	Signature on the message using private key k^-
SK_r	Symmetric key valid for encryption inside mix-zone r
t_i	Fresh timestamp
Tr_v	Intended trajectory of vehicle v

RSUs proactively include pseudonyms in CF updates and maintain a pool of these *ready-to-use* credentials for chaff message signature generation. We evaluate the size of CFs in Section VII.

B. Protocols and Services

The protocols required to implement our scheme are outlined in this section. Table I provides a summary and description of the functions and notions used.

1) Credential Acquisition: All OBUs and RSUs are assumed to have received long term credentials LK^+, LK^- after being registered at the LTCA. In order to obtain new pseudonyms, the LTCA provides a requesting entity with an authorizing ticket which is sent to the PCA [6], [11]. Subsequently, the PCA responds with a set of pseudonyms. The LTCA only issues one ticket per OBU while RSUs can obtain multiple tickets because RSUs need to generate chaff CAMs to confuse an attacker. An RSU cannot reuse a credential, otherwise an attacker would be able to classify it as a chaff pseudonym and discard the associated CAMs. On-demand pseudonym acquisition is feasible through the deployment of a highly-available and dynamically scaleable VPKI system [35].

2) Chaff Notification: Chaff messages generated by an RSU r are limited to the road segments connecting it to adjacent mix-zones with RSUs o. Consequently, r's chaff pseudonyms can only be encountered in these locations. For example, vehicles in mix-zone B in Fig. 1 can encounter chaff pseudonyms used by the mix-zones A and D. In order to not interfere with safety applications, these chaff pseudonyms are included in CFs provided to real vehicles in B's vicinity. The protocol in Fig. 2 is used by RSU r to notify its immediate neighboring RSUs o by sending the fingerprints of pseudonyms $C_{PK^+_{chaff}}$ it intends to use (step 2.1, i.e. step 1 in Fig. 2). The receiving RSUs add these fingerprints to their local CF_o which is going to be distributed to vehicles within the RSUs' vicinity. Since vehicles in transit between any o and r might be temporarily without RSU coverage, this update is subject to delays. The RSUs o use m to indicate the number of vehicles that have left towards r since the last notification (step 2.2). If $m = 0$,

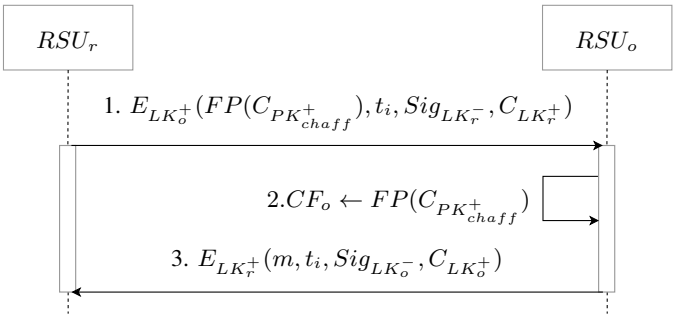


Fig. 2: RSU r notifies its adjacent RSUs o about the fingerprint of a chaff pseudonym. The RSUs reply with m indicating vehicles are in transit towards r since the last notification.

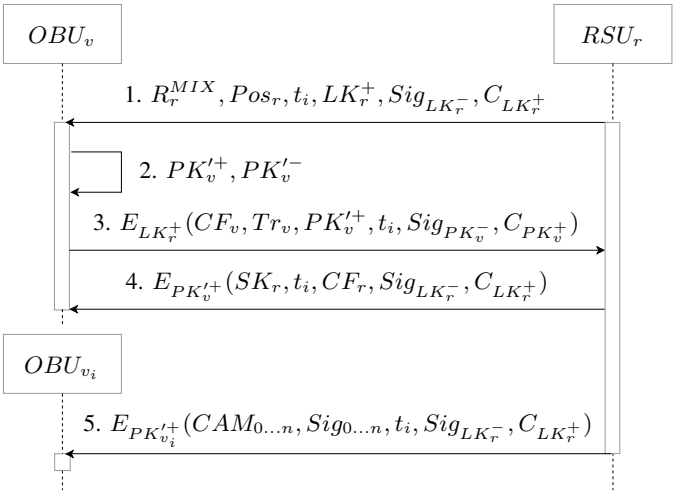


Fig. 3: Mix-zones are advertised through periodic beacons. Once inside a mix-zone, OBUs prepare a new key pair, provide their intended trajectory, and obtain the symmetric key.

no CFs need to be updated and the chaff pseudonym can be used directly. Otherwise, the chaff pseudonym cannot be used as these m vehicles would be unable to identify it as such, hence it is marked as pending. Once all relevant vehicles have received the CF update, PK^-_{chaff} is ready-to-use.

When a chaff vehicle enters a mix-zone, the fingerprint of its chaff pseudonym is removed from the CF to reduce the CF's size. In our example, chaff pseudonyms used for chaff vehicles coming from B can be removed from CF_A once they enter A. When CFs are distributed to OBUs, a signature of the originating RSU and a timestamp is included.

3) Mix-zone Participation: RSUs periodically broadcast the center Pos_r and radius R_r^{MIX} of a mix-zone (step 3.1). When an OBU entered a mix-zone, it generates a new key pair (step 3.2) later used for CAM delegation. To join a mix-zone, the approaching vehicle provides its new public key $PK_v'^+$, its current cuckoo filter CF_v, and trajectory Tr_v through that road section (step 3.3). The RSU can mark pending pseudonyms included in CF_v as ready-to-use. Encrypted with the vehicle's public key and signed by the RSU, the mix-zone specific symmetric key and local CF_r is sent back (step 3.4). Based on Tr_v, the vehicle's speed and the other real vehicles in

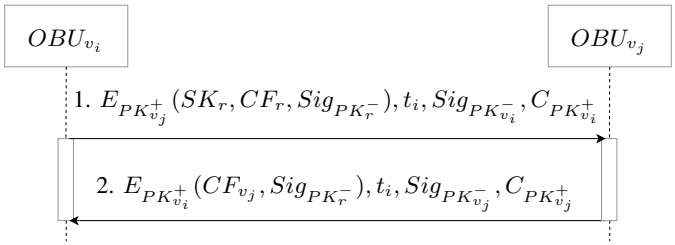


Fig. 4: OBUs forward CF updates and mix-zone keys epidemically outside RSU coverage.

the mix-zone, the RSU generates the required chaff traces. The resulting CAMs are signed using ready-to-used chaff pseudonyms. Subsequently, RSU r uses its knowledge about the trajectory of all real vehicles v_i in the mix-zone to delegate chaff CAMs that are required to originate outside its coverage. In step 3.5, these CAMs and signatures are delegated to v_i by sending them encrypted with v_i's public key $PK'^{+}_{v_i}$. v_i is going to broadcast these CAMs based on their timestamp.

4) Cuckoo Filter Update: In order to reduce the duration until a pseudonym is ready-to-use, we use epidemic propagation to disseminate CF updates between vehicles (see Fig. 4). When a vehicle v_i encounters a new peer v_j, it initiates the forwarding protocol by sending CF_{v_i} (step 4.1). After OBU v_j responds with its CF_{v_j} from the same mix-zone (step 4.2), both OBUs compare the timestamps of the CF they received. Once the signature attached to the CFs is verified, v_i and v_j update their CFs to the more recent one. Since vehicles located outside the range of a given mix-zone's RSU r are unaware of SK_r, we piggyback SK_r on this exchange.

V. Security Analysis

Messages are digitally signed under the private key of the currently valid pseudonym of an OBU or the LTC of the RSUs. They also include timestamps and hence protocol interactions are protected against injection and replay attacks.

1) Chaff Notifications: RSUs provide fingerprints of chaff pseudonyms to adjacent mix-zones to ensure safety application functionality. In the hands of an attacker, these fingerprints allow him to discard chaff messages and degrade the privacy protection offered by our scheme. Equally, if an attacker manages to inject fingerprints of real pseudonyms, OBUs would ignore real messages jeopardizing passenger safety. Hence, two RSUs exchange chaff notifications encrypted under the receiver's public key and provide a signature for integrity. The number of real vehicles in transit between two mix-zones (m) is also to be encrypted, otherwise it allows an attacker to reason about the number of chaff vehicles. For example, given $m = 0$, an attacker would conclude that no real vehicles are present and observed messages originate from chaff vehicles.

2) Mix-zone Interaction: Four types of information are transmitted when a vehicle interacts with a mix-zone's RSU: (1) the vehicle's new public key, (2) its intended route, (3) the vehicle's and RSU's CFs, and (4) chaff CAM delegations. If in possession of (1) or (2), it is trivial for an attacker to link old and new pseudonym of a vehicle. (3) and (4) enable the attacker to identify chaff pseudonyms and discard chaff CAMs. An OBU encrypts (1), (2), and (3) with the RSU's public key to initiate the protocol. The response providing the mix-zone's symmetric key and the RSU's CF is encrypted with the vehicle's public key. To protect CAM delegation (4), the RSU uses the target vehicle's public key to encrypt the communication.

3) CF Update Propagation: Our scheme epidemically propagates CF updates and mix-zone keys. A CF obtained by an attacker would enable the exclusion of all messages signed with the included chaff pseudonyms and hence defeat our scheme. A leaked mix-zone key would turn an external adversary into an internal one. To prevent an eavesdropper from acquiring any of the two pieces of information, OBUs encrypt their communication using the respective receiver's public key. The injection of forged CFs is mitigated by attaching a signature of the originating RSU. Since CFs have a fixed size (see Sec. VII) and are encrypted, an attacker cannot infer the number of active chaff vehicles from them.

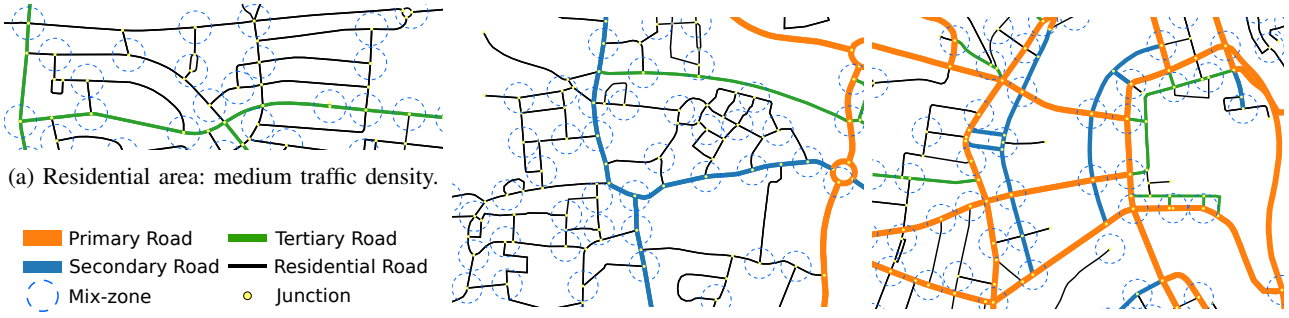
TABLE II: Metric parameters summary and description.

E	Edges $E = (n_i, n_j); n_{i,j} \in N$
$l(e)$	Length of $e \in E$
N	Junctions of the road network
M	Mix-zones $M \subseteq N$
$p_v(n)$	Probability to link v's pseudonyms when traversing $n \in M$
T_v	Trip of vehicle v, $T_v = (n_0, ..., n_i); n_i \in N$
$\|T\|$	Total length of trip T

VI. Privacy Analysis and Evaluation

Our evaluation assesses the privacy protection achieved under three scenarios with different traffic densities (see Fig. 5). These were chosen due to their range in road types and differences in driver population. We compare our chaff-based approach to the original CMIX [19] scheme that did not consider sparse vehicle availability, hereafter termed the *baseline* scheme.

a) Exposure Metric: In previous work, metrics capturing a vehicle's privacy, such as anonymity set size and tracking entropy, were used [19], [30], [33]. These assume vehicles are dispersed according to a Gaussian distribution and their mix-zone transition probabilities are normally distributed, which limits their expressiveness under real traffic conditions [38]. We define an exposure metric based on a vehicle's route length and the number of mix-zones traversed during a trip. It captures the impact of a linked change from pseudonym p to a new pseudonym p' by weighing it with the distance the vehicle will remain using p'. Intuitively, if the vehicle changes to yet another pseudonym p'' after a short period, the impact of a linking p to p' is smaller than if p' was used for a longer period as a larger piece of the journey can be attributed to the same vehicle. In Table II, we provide the parameters of a road network which we model as a graph of nodes (junctions) N and edges (roads) E. A subset M of nodes denotes the mix-zones of the network. Equation 1 shows how the privacy degradation $P(T_v)$ for vehicle v during a trip T is computed.



(a) Residential area: medium traffic density.

(b) Suburban region: low traffic density.

(c) Central region: high traffic density.

Fig. 5: The areas used for the simulation expose individual mobility patterns due to driver population and road layout. Road capacities are color coded, yellow points show junctions while dashed circles indicate encrypted mix-zones with $R^{MIX} = 50$m.

TABLE III: Region simulation parameters ($R^{MIX} = 50$m).

Parameter	Suburban	Residential	Central
Area	1.31 km^2	0.61 km^2	1.39 km^2
Junctions	69	34	61
Mix-zones	31	18	28
Average number of vehicles per mix-zone	1825	4631	6500

$$P(T_v) = \frac{\sum_{i=0}^{m} l(e_i) * p_v(n_i)}{|T_v|} \quad (1)$$

When traversing node n_i, the length of the exiting edge $l(e_i)$ is weighted by the attacker's probability p_v of successfully linking v's pseudonyms across the transition. In case $n_i \notin M$, this probability is set to 1.0, otherwise it is computed depending on the applied scheme. For the baseline scheme, traffic flow features are used, such as transition duration, transition probability, and the number of vehicles in the mix-zone [19]. For our chaff-based scheme, the number of vehicles and the generated chaff messages form the basis of the probability computation. Finally, the total length of a trip $|T_v|$ is used to normalize the exposure for vehicles with varying trip lengths.

b) Quantitative Analysis: For our simulations, we extended the implementation of the PREXT project [14] based on the Omnet++ and the Veins framework. We used the SUMO traffic simulator to produce realistic mobility traces according to the LuST dataset [21] for vehicle trajectories in the city of Luxembourg. The characterizing features of the three selected regions are provided in Table III. We show the area, the number of junctions, the number of mix-zones, and the average traffic density. Since an optimal mix-zone placement is NP-complete [39], we used a heuristic-based approach which maximizes the number of mix-zones across a map for a given encryption radius. For our initial comparison, we chose $R^{MIX} = 50$ meters. The resulting mix-zones for each region are highlighted with dashed circles in Fig. 5. Mix-zones are placed such that their encrypted areas do not overlap, otherwise, they simply merge into one larger area. We compared three different attack strategies, choosing mix-zones randomly, by traffic density, and by the number of roads exiting a mix-zone. Since selecting mix-zones by traffic density yielded the best results for the attacker, we evaluated this strategy.

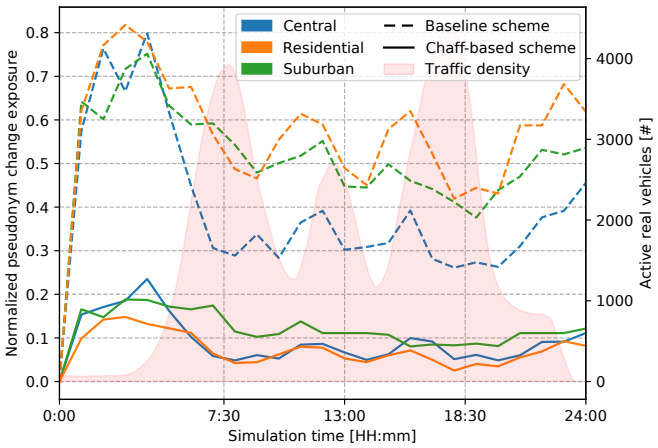


Fig. 6: Privacy degradation comparison: lines show exposure values for the 90$^{\text{th}}$ percentile of the vehicle population.

After simulating both schemes for the three scenarios over a full-day period, we obtained the results depicted in Fig. 6. They provide the exposure values for the 90$^{\text{th}}$ percentile of the vehicle population. We found that the privacy exposure is inversely correlated with real vehicle density. This can be seen during the rush hours around 7:30, 13:00, and 18:30 which coincide with a reduction in exposure values. Due to very low traffic density from midnight to 5:00 am, this period provides vehicles the least privacy protection. Regional variations, e.g., between the central and residential area, also contribute to differences in privacy exposure levels. In contrast, our scheme appears more resilient against these fluctuations providing a homogeneous level of privacy protection.

Next, we evaluated the effect of different encryption radii on the performance of both schemes. As shown in Fig. 7, we conducted simulations for $R^{MIX} = [50, 100, 150, 200, 250]$. We considered an attacker who has several observation spots with limited coverage [31] and a global attacker, capable of covering the entire network, in the residential and central region. These regions were chosen since they exhibited the best

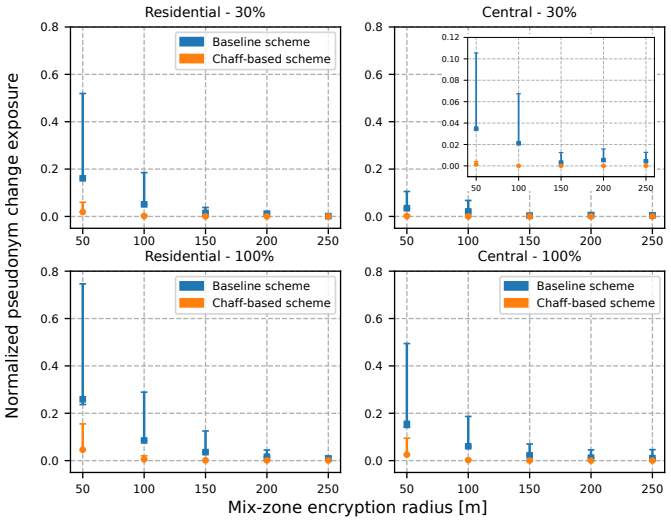


Fig. 7: Effect of mix-zone encryption range on vehicle exposure. The attackers observe 30% and 100% of the mix-zones.

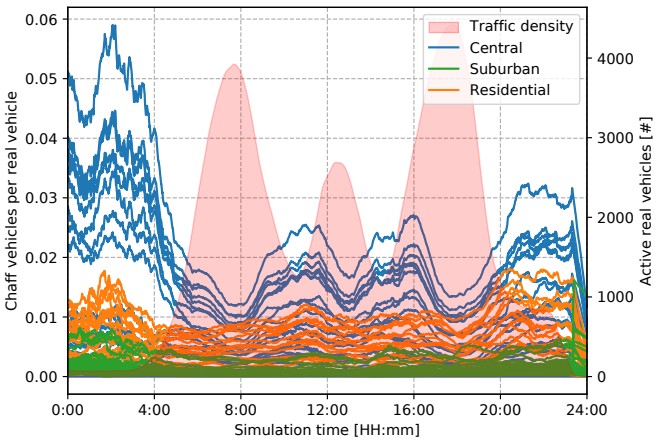


Fig. 8: Active chaff vehicles per real vehicle. The shaded area denotes number of real vehicles. One line per mix-zone: 28, 31, 18 for the central, suburban, and residential region.

TABLE IV: Different encryption radii affect the number of simultaneously active chaff pseudonyms, chaff message generation, and pseudonym pool size.

Mix-zone encryption radius [m]	50	100	150	200	250
Max. Active chaff pseudonyms [#]	68	176	165	99	66
Max. CAM generation $[msg/s]$	240	848	1321	831	634
Max. Pseudonym pool size [#]	5	4	3	4	3

and worst privacy protection in our analysis with $R^{MIX} = 50$. We found that user privacy improves with increasing encryption range; however, the baseline scheme converges slower than our chaff-based scheme and exposes larger variations, confirming its dependence on mobility patterns. After an initial improvement in privacy protection by increasing R^{MIX} from 50 to 100 meters for the chaff-based scheme, the privacy protection remains optimal thereafter. These figures indicate that while the baseline scheme offers more user privacy with a larger R^{MIX}, a chaff-based scheme can operate almost independent of the value chosen for R^{MIX}. A smaller R^{MIX} means that a vehicle enters and leaves more mix-zones, consequently its journey is divided into more segments which offers more privacy protection according to Freudiger et al. [19].

VII. Performance Evaluation

We evaluate the performance of our system using three metrics: (1) chaff pseudonym pool size, (2) the number of simultaneously active chaff pseudonyms, and (3) the RSU signature generation overhead. Considering different values for R^{MIX}, the remainder of this section discusses resource requirements on the basis of these metrics. The results acquired over the whole 24h simulation are shown in Table IV.

For the number of active chaff vehicles, we are interested in two aspects. Firstly, whether real vehicles can alleviate the number of required chaff vehicles, and secondly the absolute computation overhead experienced by an RSU for chaff CAM signature generation. Given the influx of vehicles into a mix-zone under consideration of the availability of real vehicles, we can compute the number of required chaff vehicles. Intuitively, this number should be inversely correlated with the traffic density. Figure 8 shows each mix-zone's active chaff pseudonyms per real vehicle. As expected, we found that with an increase in real vehicle density during peak hours (around 7:30, 13:00 and 17:30), less chaff vehicles are required to achieve the same level of privacy protection. Secondly, based on trace length and the number of required chaff vehicles, we can compute the rate with which RSUs have to sign chaff messages. Studies on Nexcom boxes (Dual-core 1.66 GHz, 1GB memory) from the PRESERVE project show that the average latency to generate Elliptic Curve Digital Signature Algorithm (ECDSA) signatures, over 256 bit prime fields, is 0.4 ms [40]. The results indicate that even modest computing resources can support generating 6742 signatures per second. Note that RSUs will be provided with higher computational power; thus, generating thousands of chaff CAMs will not incur excessive workload on the RSUs.

In order to ensure the timely generation of chaff messages, RSUs retain a pool of ready-to-use pseudonyms. We provide an estimate for the size of this pool, by measuring the maximum propagation delay between adjacent mix-zones. Using the arriving real vehicles and hence the number of chaff vehicles, we can compute the required chaff pseudonyms. According to our results, a pool size of five chaff pseudonyms ensures the timely availability of ready-to-use pseudonyms.

For safety applications to operate correctly, CFs are distributed to vehicles in transit between adjacent RSUs. We provide an upper bound for their size to assess the storage and communication cost. After a full-day simulation, we found a maximum of 175 and 160 simultaneously active chaff pseudonyms across all mix-zone for $R^{MIX} = [100, 150]$. After examining the mix-zone placement, we can explain this peak through a lower mix-zone density. Given the constraint of non-overlapping zones, mix-zones are placed sparser across the same area causing longer unencrypted segments between

zones. The results are longer chaff traces and hence simultaneously active chaff pseudonyms. Even though a CF containing these pseudonyms is sufficient for safety applications, it requires to be continuously updated, leading to an increased overhead during epidemic propagation. As mitigation, a larger number of chaff pseudonyms can be proactively included in the CF. For example, using the maximum simultaneous active chaff pseudonyms, i.e., 176, we can compute that 316,800 chaff pseudonyms are sufficient to provide a CF that needs to be updated every 30 minutes. This amounts for a CF with 3.63 MB in size at a false positive rate of $p = 10^{-20}$ [37]. Given a data rate of several Mbit/s for modern IEEE 802.11p interfaces [1], this means CF updates do not pose a significant communication overhead.

VIII. Conclusion and Future Work

We proposed a novel scheme for private pseudonym change that offers a constant level of privacy independent of the operation area, mix-zone encryption radius, time of the day, and vehicle arrival rates. Our scheme ensures that safety applications remain unaffected by the introduction of chaff vehicles. Through an extensive analysis, we showed the security and privacy properties of our system and analyzed the performance in comparison to a previously proposed CMIX scheme. Without significant additional overhead, our scheme enhanced user privacy up to 76%, independent of the underlying traffic conditions. As future work, we plan to investigate how parameters such as encryption radius, number of mix-zones, and ephemeral mix-zone encryption keys can provide resilience against an internal attacker, e.g., malicious OBUs.

Acknowledgements

This work has been partially supported by the Swedish Foundation for Strategic Research (SSF) SURPRISE project and the British Engineering and Physical Sciences Research Council (EPSRC).

References

[1] "IEEE Standard for Wireless Access in Vehicular Environments (WAVE) –Networking Services," *IEEE Vehicular Technology Society*, Jan. 2016.

[2] ETSI, "Intelligent Transport Systems (ITS); Vehicular Communications; Basic Set of Applications; Definitions," TR-102-638, Jun. 2009.

[3] P. Papadimitratos and et al, "Securing Vehicular Communications-Assumptions, Requirements, and Principles," in *ESCAR*, Berlin, Germany, Nov. 2006.

[4] P. Papadimitratos *et al.*, "Secure Vehicular Communication Systems: Design and Architecture," *IEEE Communications Magazine*, vol. 46, no. 11, pp. 100–109, Nov. 2008.

[5] M. Khodaei, H. Jin, and P. Papadimitratos, "Towards Deploying a Scalable & Robust Vehicular Identity and Credential Management Infrastructure," in *IEEE VNC*, Paderborn, Germany, Dec. 2014.

[6] M. Khodaei *et al.*, "SECMACE: Scalable and Robust Identity and Credential Management Infrastructure in Vehicular Communication Systems," *IEEE T-ITS*, vol. 19, no. 5, pp. 1430–1444, May 2018.

[7] B. Wiedersheim *et al.*, "Privacy in Inter-vehicular Networks: Why Simple Pseudonym Change is not Enough," in *International Conference on WONS*, Kranjska Gora, Slovenia, Feb. 2010, pp. 176–183.

[8] L. Buttyán *et al.*, "SLOW: A Practical Pseudonym Changing Scheme for Location Privacy in VANETs," *IEEE VNC*, Oct. 2009.

[9] M. Khodaei *et al.*, "POSTER: Privacy Preservation through Uniformity," in *ACM WiSec*, Stockholm, Sweden, June 2018, pp. 279–280.

[10] K. Emara *et al.*, "Vehicle Tracking using Vehicular Network Beacons," in *IEEE WoWMoM*, Madrid, Spain, June 2013.

[11] M. Khodaei and P. Papadimitratos, "Evaluating On-demand Pseudonym Acquisition Policies in Vehicular Communication Systems," in *ACM IoV/VoI*, Paderborn, Germany, July 2016.

[12] P. Golle and K. Partridge, "On the Anonymity of Home/Work Location Pairs," in *Pervasive computing*. Springer, 2009, pp. 390–397.

[13] M. Gerlach and F. Guttler, "Privacy in VANETs using Changing Pseudonyms - Ideal and Real," in *IEEE VTC*, Dublin, Ireland, Apr. 2007.

[14] K. Emara, "Poster: PREXT: Privacy Extension for Veins VANET Simulator," in *IEEE VNC*, Columbus, OH, USA, Dec 2016. [Online]. Available: tiny.cc/5oi6xy

[15] K. Emara *et al.*, "CAPS: Context-Aware Privacy Scheme for VANET Safety Applications," in *ACM WiSec*, New York, NY, USA, June 2015.

[16] K. Sampigethaya *et al.*, "AMOEBA: Robust Location Privacy Scheme for VANET," *IEEE JSAC*, vol. 25, no. 8, pp. 1569–1589, Oct. 2007.

[17] S. Lefèvre *et al.*, "Impact of V2X Privacy Strategies on Intersection Collision Avoidance Systems," in *IEEE VNC*, Boston, MA, Dec. 2013.

[18] G. P. Corser *et al.*, "Effect on Vehicle Safety of Nonexistent or Silenced Basic Safety Messages," in *ICNC*, Kauai, HI, USA, Feb. 2016, pp. 1–5.

[19] J. Freudiger *et al.*, "Mix-zones for Location Privacy in Vehicular Networks," in *Win-ITS*, Vancouver, BC, Canada, Aug. 2007.

[20] X. Liu *et al.*, "Traffic-aware Multiple Mix-zone Placement for Protecting Location Privacy," in *IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012.

[21] L. Codeca, R. Frank, and T. Engel, "Luxembourg Sumo Traffic (LuST) Scenario: 24 Hours of Mobility for Vehicular Networking Research," in *IEEE VNC*, Kyoto, Japan, Dec. 2015.

[22] B. Palanisamy and L. Liu, "Attack-Resilient Mix-zones over Road Networks: Architecture and Algorithms," *IEEE TMC*, vol. 14, no. 3, pp. 495–508, 2015.

[23] J. Krumm, "Inference Attacks on Location Tracks," in *International Conference on Pervasive Computing*, Toronto, Canada, May 2007.

[24] C. Vaas, P. Papadimitratos, and I. Martinovic, "Poster: Increasing Mix-Zone Efficacy for Pseudonym Change in VANETs Using Chaff Messages," in *ACM WiSec*, Stockholm, Sweden, June 2018.

[25] J. Bellatti *et al.*, "Driving Habits Data: Location Privacy Implications and Solutions," in *IEEE S&P*, vol. 38, no. 1, Jan. 2017, pp. 12–20.

[26] S. Narain, T. D. Vo-Huu, K. Block, and G. Noubir, "Inferring User Routes and Locations Using Zero-Permission Mobile Sensors," in *IEEE S&P*, San Jose, CA, USA, May 2016, pp. 397–413.

[27] X. Gao *et al.*, "Elastic Pathing: Your Speed is Enough to Track You," in *ACM UbiComp*, Seattle, Washington, Sep. 2014, pp. 975–986.

[28] Q. Technologies, "Leading the World to 5G: Cellular Vehicle-to-Everything (C-V2X) Technologies," tiny.cc/u4v2xy, Jun. 2016.

[29] L. Huang *et al.*, "Enhancing Wireless Location Privacy using Silent Period," in *IEEE WCNC*, New Orleans, LA, USA, Mar. 2005.

[30] K. Sampigethaya *et al.*, "CARAVAN: Providing location privacy for VANET," in *Embedded Security in Cars*, Cologne, Germany, Nov. 2005.

[31] D. Förster *et al.*, "An Evaluation of Pseudonym Changes for Vehicular Networks in Large-Scale, Realistic Traffic Scenarios," *IEEE Transactions on ITS*, pp. 1–6, Dec. 2017.

[32] A. Tomandl, F. Scheuer, and H. Federrath, "Simulation-based Evaluation of Techniques for Privacy Protection in VANETs," in *IEEE WiMob*, Barcelona, Spain, Oct. 2012, pp. 165–172.

[33] A. Wasef and X. Shen, "REP: Location privacy for VANETs using random encryption periods," *Mobile Networks and Applications*, vol. 15, no. 1, pp. 172–185, Feb. 2010.

[34] M. Khodaei and P. Papadimitratos, "The Key to Intelligent Transportation: Identity and Credential Management in Vehicular Communication Systems," *IEEE VT Magazine*, vol. 10, no. 4, pp. 63–69, Dec. 2015.

[35] H. Noroozi, M. Khodaei, and P. Papadimitratos, "DEMO: VPKIaaS: A Highly-Available and Dynamically-Scalable Vehicular Public-Key Infrastructure," in *ACM WiSec*, Stockholm, Sweden, June 2018.

[36] M. Khodaei and P. Papadimitratos, "Efficient, Scalable, and Resilient Vehicle-Centric Certificate Revocation List Distribution in VANETs," in *ACM WiSec*, Stockholm, Sweden, June 2018, pp. 172–183.

[37] B. Fan *et al.*, "Cuckoo Filter: Practically Better Than Bloom," in *ACM CoNEXT*, Sydney, Australia, Dec. 2014, pp. 75–88.

[38] Y. Sun, B. Zhang, B. Zhao, X. Su, and J. Su, "Mix-zones Optimal Deployment for Protecting Location Privacy in VANET," *Peer-to-Peer Networking and Applications*, vol. 8, no. 6, pp. 1108–1121, Nov. 2015.

[39] J. Freudiger, R. Shokri, and J.-P. Hubaux, "On the Optimal Placement of Mix Zones," in *PETS*, Berlin, Heidelberg, Aug. 2009, pp. 216–234.

[40] M. Feiri *et al.*, "Preparing secure vehicle-to-x communication systems," Deliverable 3.2, FOT Trial 2 Results, 2015. [Online]. Available: https://www.preserve-project.eu/deliverables