MARP: A Distributed MAC Layer Attack Resistant Pseudonym Scheme for VANET

Zishan Liu, Student Member, IEEE, Zhenyu Liu, Student Member, IEEE, Lin Zhang, Member, IEEE, and Xiaodong Lin, Fellow, IEEE

Abstract—Modern vehicles are equipped with wireless communication technologies, allowing them to communicate with each other and forming large self-organized ad hoc networks (or vehicular ad hoc networks, VANETs). VANETs, while promising new approaches for improving road safety, require privacy of vehicles (or drivers) to be protected from a variety of threats. Although pseudonym schemes have provided a promising solution at the upper layers, privacy attacks could still be carried out from medium access control (MAC) layer. In this paper, we first introduce a new MAC layer context linking attack, which could link the old and new pseudonyms of a vehicle by analyzing its transmission characteristics. To deal with the attack, we propose a time division multiple access based MAC-layer-Attack-Resistant Pseudonym (MARP) scheme. Unlike traditional approaches that design the MAC protocols and pseudonym schemes separately, MARP allows vehicles to change their transmission slots and pseudonyms collaboratively. Thus, the unlinkability is guaranteed. Taking the pseudonym age, anonymity set size and time-to-confusion as the location privacy metrics, we derive an analytical model to quantify location privacy achieved in MARP. The analytical model is general to be applied for other pseudonym schemes. Extensive simulation results have validated the analytical model, showed that MARP can resist the MAC context linking attack and guarantee location privacy and efficient transmission for vehicles in VANETs.

Index Terms—location privacy, pseudonym, MAC, VANET

1 INTRODUCTION

Vehicular ad hoc networks (VANETs) are expected to enhance the transportation safety and efficiency, as well as provide infotainment services. Through Dedicated Short Range Communication (DSRC), vehicles can periodically broadcast “beacons” or basic safety messages (BSMs) including their locations, speed and acceleration/deceleration every 100-300ms to the surrounding vehicles and roadside units (RSUs). BSMs play a vital role in creating cooperative neighborhood awareness and facilitating the safety applications. Unfortunately, this would facilitate adversaries tracking the trajectory of vehicles and hence compromising location privacy.

It is of practical significance to protect location privacy in VANETs. Recently, many security and privacy-preserving mechanisms and protocols have been developed. For example, standardization bodies such as IEEE 1609.2 [1] and ETSI [2], have proposed the public key infrastructure (PKI) based cryptography to protect vehicular communications. To ensure the communication authenticity, integrity, identity privacy and non-repudiation, pseudonyms certificates are used with the corresponding short-lived private keys to sign the BSMs. Furthermore, the pseudonyms need to change frequently and suitably [3], so that the messages originating from the same vehicle are unlinkable, so as to protect the location privacy.

However, a dilemma between location privacy and transmission efficiency arises in VANETs. This is because the fundamental purpose of communication is to transmit data efficiently and reliably. It is crucial to have a well-designed medium access control (MAC) layer protocol for VANETs. Previous work considers the design of pseudonym scheme for location privacy, and the MAC layer protocols for transmission efficiency separately. The MAC layer transmission context could still provide a clue to link the new pseudonym with the old ones and compromise the location privacy.

Various MAC layer protocols have been proposed for VANETs. The DSRC has employed IEEE 802.11p [4] for the MAC layer operation, in which BSMs broadcasting is based on carrier sense multiple access/collision avoidance (CSMA/CA). However, prior studies [5], [6] have shown that for higher density of vehicles, IEEE 802.11p suffers a significant performance degradation and behaves similar to ALOHA. The CSMA based MAC protocol is only adaptable to low-to-moderate channel loads. Meanwhile, we found that a context-linking attack is possible to launch using the transmission cyclicity of the BSMs. Adversaries could utilize the MAC layer activity time as the “fingerprint” of vehicles to infer their identities regardless of the changing pseudonyms.

Besides the CSMA based schemes, SDMA schemes [7] have been proposed to assign channel resources according to the locations of vehicles to avoid signal interference among vehicles in proximity. Obviously, they contradict with the location privacy requirements. Many researchers prefer to TDMA based mechanisms for VANETs, such as VeMAC [8] and CFR MAC [9], etc. In TDMA, time is divided into synchronized frames that consist of a number of time slots. Each vehicle reserves one or more time slots to transmit BSMs periodically and keeps silent during other slots that have been reserved by its neighbors. TDMA schemes have been promising to fulfill the transmission requirements of the safety messages. Unfortunately, the more coordination of a MAC protocol, the more likely to facilitate the context linking attack. Adversaries could link the new pseudonym with the old ones by monitoring the slot utilization information of a targeted vehicle. In summary, we cannot design the
MAC layer protocol and pseudonym schemes separately to guarantee the location privacy as well as location privacy for VANETs.

To illustrate the MAC layer context linking attack via the transmission correlation, an example is shown in Fig. 1. The vehicles access the wireless medium to broadcast the BSMs periodically. At the first synchronized control channel interval, i.e., frame $F_1$, the vehicles of Bob, Alice and Grey use the pseudonyms $x_1$, $y_1$, and $z_1$ to broadcast BSMs at time slot $t_2$, $t_4$, and $t_6$ respectively. They switch to use the new pseudonyms $x_2$, $y_2$, and $z_2$ at the frame $F_2$. Under this circumstance, they are indistinguishable in terms of the pseudonyms at the upper layer, e.g., application layer. However, their driving trajectories can be monitored continuously based on the BSMs transmitted at the specific time instants from the MAC layer. The transmission cyclicity can be used to link the new pseudonyms with the old ones, and even correlate with the driving route and sensitive locations (home, working place, etc.) together to de-anonymize Alice, Bob and Grey.

Therefore, the pseudonym schemes should be designed collaboratively with the message transmission as they influence each other. Lefèvre et al. [10] take the road intersection as the traffic scenario to analyze the impact of the privacy strategies on intersection collision avoidance (ICA) system. They draw a conclusion that the requirements of safety applications should be taken into account when designing pseudonym changing strategies. Fonseca et al. [11] suggest that the pseudonyms must change consistently across multiple layers, i.e., the IP and MAC address must be hidden or changed simultaneously with the pseudonyms to avoid trivial linking from other identifiers. However, it still lacks a scheme to resolve the new MAC layer attack caused by the inconsistency between channel accessing and pseudonym changing. The separated design of pseudonym schemes rarely consider the communication overhead and impact on the MAC layer performance, and some are even unsuitable to work under the coordinated based MAC protocols. Therefore, the problem of guaranteeing location privacy while achieving transmission efficiency needs to be addressed collaboratively in VANETs. To overcome these challenges, we propose a MAC layer Attack Resistant Pseudonym (MARP) scheme for the BSMs broadcasting in VANETs. Specifically, the main contributions of the paper are listed as follows.

Firstly, we identify the new MAC layer context linking attack facilitated by the inconsistent operation of the pseudonym changing and periodical MAC layer activities. We analyze the attack under both CSMA and TDMA based protocols.

Secondly, we propose the MARP scheme to guarantee the location privacy and fulfill the BSM transmission requirements in VANETs. In MARP, vehicles change the pseudonyms and time slots by cooperatively negotiating with their neighbors to construct mix-zones based on the current pseudonym age. The key idea is to divide the synchronized frames (i.e., control channel interval) to a dedicated transmission period and a flexible schedule period. The dedicated transmission period consists of separated time slots used to broadcast safety messages. Vehicles randomly shuffle their transmission slots when changing pseudonyms, so as to keep unlinkable to the adversaries. The schedule period is used for the communication between vehicles and RSUs or trusted authorities (TAs), e.g., to request pseudonyms or receive the certificate revocation list (CRL) from the TA. When a mix-zone is constructed, vehicles becomes indistinguishable by releasing their transmission slots and old pseudonyms cooperatively. By this way, the unlinkability of the pseudonyms can be achieved.

Finally, the location privacy achieved in MARP is quantified using a detailed analytical model. The privacy metrics include the pseudonym age, anonymity set size and time-to-confusion. Based on the mean field theory, the analytical model presents a general solution for the location privacy measurement regardless of the transmission pattern and vehicle distribution when the vehicle number is large in the network. Extensive simulation results corroborate the analytical model, and verify that MARP can resist the new MAC layer attack. It shows that MARP is efficient as it introduces negligible communication overhead and bounded transmission delay. The scheme fills the blank of the cross-layer design of pseudonym schemes.

The remainder of the paper is organized as follows. Section 2 reviews the related work. Section 3 presents the preliminaries of the paper and the MAC layer context linking attack on the location privacy in VANETs. The MARP scheme is proposed in Section 4. Subsequently, we describe the analytical model to evaluate the location privacy achieved in MARP in Section 5. The analytical results are provided and verified by the simulation results in Section 6. The performance of MARP is also compared with existing schemes to show that MARP achieves provable location privacy with sustainable sacrifice on the transmission reliability. Finally, Section 7 concludes the paper.

2 RELATED WORK

Due to the broadcast nature of VANETs, V2V and V2R communications are vulnerable to the long-term and large scale tracking from adversaries. Thus, the identities and locations that reveal and link vehicles or drivers must be protected to achieve the location privacy [12]. Pseudonymous signature based schemes are applied to enable the authentication, message integrity, non-repudiation while preserving the anonymity for each vehicle. Currently, the most relevant pseudonym schemes for VANETs have been reviewed by Petit et al. [13]. They point out that the lifecycle of an efficient pseudonym scheme includes the pseudonym issuance, use, change, resolution and revocation. Vehicles usually request a pool of pseudonyms from the pseudonym
certificate authorities (PCAs) during the traveling [1], [2], [14]. Each pseudonym has a minimum lifetime to ensure the stable transmission and keep a moderate cost of the communication between the vehicles and the PCAs. Meanwhile, the short-lived pseudonyms have to change frequently so as to break the temporal and spatial link-ability of the vehicles. The ETSI standard [2] recommends that pseudonyms should be changed every 5 minutes, while the SAE J2735 suggests the value as 120s or 1 km [15]. However, simple pseudonym changing alone or in sparse scenarios is still not sufficient to confuse an adversary.

The radio silent technique has been proposed by [16] to provide unobservable mix-context for vehicles when they merging or changing lanes. The mix-zone concept [17] is also widely applied to construct appropriate areas by a number of vehicles to change their pseudonyms simultaneously. A cryptographic mix zones (CMIX) is introduced by Freudiger et al. [18] which allows vehicles to obtain a symmetric key from the RSU and use it to encrypt all the messages while they driving within the mix-zone. Lu et al. [3] suggest to assign the social spots such as intersections and parking lots as mix-zones to gather more vehicles to change pseudonyms simultaneously.

Fixed mix-zone based schemes present inefficiency and unpractical drawbacks, especially when the traffic density is low [19]. Pseudonym changing should be dynamically according to the traffic density and speeds. Yu et al. [20] propose an optimal pseudonym changing scheme called MixGroup, in which vehicles can negotiate with the encountered vehicles and exchange their pseudonyms with each other. The key idea of the work is based on the observation that the traffic may be sparse that only a few vehicles meet in global social spots, while each individual has many opportunities to meet other vehicles during the traveling. Therefore, by accumulatively exploiting the sparse meeting chances and integrating the group signature, MixGroup can enlarge the uncertainty of pseudonym mixture. However, MixGroup requires several rounds of communication between any two vehicles when they negotiate to exchange pseudonyms. Wei et al. [21] propose to use the obfuscating position, speed and heading of vehicle within the radius of the safe distance calculated by a safety analysis algorithm to preserve location privacy with moderate effect on the traffic safety. Meanwhile, they proposed the dynamical adjustment of the silent period based on the distance of vehicles. To guarantee the traffic safety, the closer the vehicles driving to each other, the shorter the silent period. To protect location privacy without impacting the traffic safety in VANETs, the work [22] proposes a traffic aware pseudonym changing strategy that suggests vehicles construct mix-zones using the radio silent technique when encountering traffic congestions. During traffic congestion, vehicles are running with low speeds and high probability to change lanes just before entering the traffic congestion. However, the strategy requires high communication overhead to detect the start and end of congestion.

Whyte et al. [14] present the Security Credential Management System (SCMS) for V2V communications, which has been one of the leading candidate design for the V2V security backend. In SCMS, vehicles are issued frequent changing pseudonym certificates (e.g., every 5 minutes), and the provisioning of those certificates is divided among multiple organizations. By this way, the privacy against attacks from SCMS insiders can be increased. However, the authors also point out that the BSMs may potentially be tracked in more intelligent ways that are not addressed by the SCMS. In [23], the authors propose a vehicular PKI based system, namely SECMACE to enhance the user privacy by preventing linking pseudonyms based on the pseudonym timing information. However, the new MAC transmission correlated attack can still be launched to compromise the location privacy.

Privacy metrics, on the other hand, are of significant importance to evaluate and measure the location privacy quality. Typically, the anonymity set size [3], [18], entropy [20], [24], tracking probability [24], [25], [26], pseudonym age [26] and time-to-confusion [27] are usually applied to evaluate and guide the pseudonym changing strategies in VANETs. The anonymity set size measures the number of vehicles in an area that are indistinguishable from each other to an adversary [28]. In [29], a detailed analytical model for the random pseudonym change scheme is proposed to calculate the anonymity set size given the age of pseudonyms. The entropy is proposed based on the Shannon’s classic measure for an anonymity set that has the maximal value when all the vehicles of the set are equally likely to be the one of target to the adversary [30]. The tracking probability is calculated as the probability that an adversary can link the new and old pseudonym of a specific vehicle in an anonymity set. Pseudonym age can be used to decide whether and when to change the pseudonym. To evaluate the age of pseudonyms, Freudiger et al. [26] develop a framework upon the vehicle cooperation probability, traffic mobility, pseudonym cost and the aging rate. As the quality of privacy or the degree of privacy risk strongly depends on the maximum duration adversaries can track a vehicle, time-to-confusion is thereby proposed [27]. It measures the duration an adversary could correctly follow the trace until it could not determine the next sample of a vehicle with sufficient certainty.

Intuitively, the time-to-confusion is limited by the pseudonym age if the mix-zones bring enough confusion to the adversary. However, it is not always the case since more intelligent attacks have surfaced. Wernke et al. [31] point out that an adversary can jeopardize the unlink-ability of pseudonyms using the context linking attacks in VANETs. Emara et al. [32] describe a method to track vehicles based on beacon messages in VANETs. In [33], Bloessl et al. present a scrambler attack on location privacy based on the physical characteristics of the vehicles.

In this paper, the MARP scheme is proposed to deal with the MAC layer context linking attack in VANETs. The attack jeopardizes the unlink-ability and increases the tracking duration based on the targeted vehicle’s transmission activities. To resist the attack, MARP designs the key idea to randomly permute their transmission time when vehicles changing pseudonyms.

3 Preliminaries

In this section, we firstly introduce the assumptions made throughout the paper. A list of notations used in this paper
is shown in Table 1.

<table>
<thead>
<tr>
<th>Notations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_i$</td>
<td>A representative vehicle</td>
</tr>
<tr>
<td>$R$</td>
<td>The transmission range of a vehicle</td>
</tr>
<tr>
<td>$U_t$</td>
<td>The tracking uncertainty of $v_i$ by an adversary</td>
</tr>
<tr>
<td>$PID_{i,j}$</td>
<td>The jth pseudonym of $v_i$ signed by the PCA</td>
</tr>
<tr>
<td>$(K_{i,j}, k_{i,j})$</td>
<td>The pseudonymous key pair of $v_i$ corresponding to the jth pseudonym</td>
</tr>
<tr>
<td>$Cert_{i,j}$</td>
<td>The pseudonym certificate of $v_i$ corresponding to the jth pseudonym</td>
</tr>
<tr>
<td>$Sign_{i,j}(M)$</td>
<td>The signature of message M signed by $v_i$ using the $k_{i,j}$</td>
</tr>
<tr>
<td>$N_{ts}$</td>
<td>The number of time slots in a frame</td>
</tr>
<tr>
<td>$N$</td>
<td>The total number of vehicles</td>
</tr>
<tr>
<td>$C(v_i)$</td>
<td>The cooperative neighbor set of $v_i$</td>
</tr>
<tr>
<td>$ts_i$</td>
<td>The transmission slot reserved by $v_i$ in the DT</td>
</tr>
<tr>
<td>$s_i$</td>
<td>The slot status of $ts_i$</td>
</tr>
<tr>
<td>$\tau$</td>
<td>The cooperation pseudonym age threshold</td>
</tr>
<tr>
<td>$\varepsilon_p$</td>
<td>The pseudonym age</td>
</tr>
<tr>
<td>$M(z,t)$</td>
<td>The occupancy measure of the vehicles with the pseudonym age equal to $z$ at time $t$</td>
</tr>
<tr>
<td>$f(z,t)$</td>
<td>PDF of pseudonym age equals to $z$ at time $t$</td>
</tr>
<tr>
<td>$F(z,t)$</td>
<td>CDF of pseudonym age at time $t$</td>
</tr>
<tr>
<td>$p(c(t))$</td>
<td>Probability of at least one neighbor cooperates at time $t$</td>
</tr>
<tr>
<td>$c(t)$</td>
<td>Probability of any neighbor cooperates at time $t$</td>
</tr>
<tr>
<td>$t_c$</td>
<td>The time-to-confusion of an adversary</td>
</tr>
<tr>
<td>$n_c$</td>
<td>The average anonymity set size in a mix-zone</td>
</tr>
<tr>
<td>$ts_{min}$</td>
<td>The minimum silent period</td>
</tr>
<tr>
<td>$ts_{max}$</td>
<td>The maximum silent period</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>The average arrival rate of vehicles</td>
</tr>
<tr>
<td>$\rho$</td>
<td>The average traffic density on the roads</td>
</tr>
</tbody>
</table>

### 3.1 Network Model

In this paper, we design the protocol based on the certificate management architecture SCMS, proposed by Whyte et al. [14]. The SCMS has defined the duties of multiple entities in the overall architecture. In SCMS, the root/enrollment certificate authority (CA) is to issue the enrollment certificates that act as passport for each vehicle to request further pseudonym certificates. For any vehicle, before entering the network, the bootstrap and enrollment is required for the on-board units (OBUs) to obtain the certificates of the CAs, the enrollment certificate and the information to locate the registration authorities. Before a new trip, vehicles need to request new pseudonym certificates. The requests are validated, processed and then forwarded by the registration authority to the pseudonym certificate authority (PCA). The PCA is responsible to issue short term pseudonym certificates to devices and the request coordination entity is to ensure that a device does not request more than one set of certificates for a given time period. We omit the description of the whole framework but refer the readers to paper [14] for a more detailed study. In this paper, the certificate management related entities are collectively referred to as the TAs. The considered VANET system mainly consists of three entities: the OBUs equipped in the vehicles, the RSUs along the roads and the TAs.

- Each vehicle is equipped with an OBU to broadcast the BSMs. The precise driving information and time synchronization are achieved by the global positioning system (GPS) in each vehicle. The transmission power level is assumed to be fixed and known to each other. All the physical channels are symmetric and the communication range is denoted as $R$.
- RSUs are responsible for communicating as the gateways to deliver some certificate materials such as the certificate revocation list (CRL) from the TA to OBUs and collecting the traffic information from the vehicles within its coverage. Assume RSUs are not trusted and they access the channel and authenticate messages in the similar approach as vehicles.
- The TAs are responsible for the device bootstrap, pseudonym certificate provision and revocation of malicious entities. Before joining the system, each vehicle and RSU need to register with the TAs. As general, we assume the TAs have the highest security level and are infeasible to be compromised. The TAs may be curious to compromise the location privacy of vehicles, therefore their duties are unlinkable for the pseudonym provisioning.

### 3.2 Threat Model

The Global Passive Adversary (GPA) model [13] is assumed in this paper, which aims to track any vehicle in a region-interest by eavesdropping rather than compromising its transmission. In this paper, we assume the main motivation of GPA is to track the locations and driving paths of the vehicles, and thereby explore their lifestyles and sensitive information. In this paper, the tracking techniques released by the adversary mainly include: 1) collect beacon messages and context information (e.g., MAC address, coordination information or the pseudonyms contained in the messages) to increase the link-ability of the target vehicle. 2) reconstruct the driving paths from the location samples of the targeted vehicles. In this paper, we ignore the vision-based and the physical layer based attacks, and only consider the radio communication based GPA.

### 3.3 The MAC Layer Context-Linking Attack

Previous work considers the MAC layer protocols and pseudonym schemes separately. Unfortunately, their inconsistent operations facilitate the context-linking attack, by which the GPA can easily track a specific vehicle by analyzing its transmission periodicity of the messages. To formally describe the attack, we start by considering the tracking uncertainty of an adversary in a mix-zone.

The adversary $A$ observes a set of $n$ vehicles change pseudonyms simultaneously in a mix-zone at the time instant $T$. $A$ compares the new set of pseudonyms with the old pseudonyms, and based on the mobility and other context information to predict the most probable linking.
adversary’s uncertainty on the linking of a representative vehicle \( v_i \) is denoted as \( U_i \) and computed as follows [25].

\[
U_i(T) = - \sum_{j=1}^{n} p_j \log_2(p_j)
\]

(1)

where \( \sum_{j=1}^{n} p_j = 1 \), and \( p_j \) measures the probability that the new pseudonym \( PID'_i \) of vehicle \( v_i \) is linked with the old pseudonym \( PID_i \) of \( v_i \). \( U_i \) is upper-bounded by \( \log_2 n \), when there is no side information to distinguish \( v_i \) from the other vehicles in the anonymity set, i.e., \( p_j = 1/n \). However, the successful linking probability \( p_j \) could be highly increased and even become 1 when \( A \) has the side information of \( v_i \), e.g., the MAC layer context. As a consequence, the vehicle loses privacy entropy and its privacy level is decreased.

When vehicles exploit the CSMA based protocol for the MAC layer operation, we assume that the BSM generation pattern of vehicles is periodical and independent from each other. As shown in Fig. 2, \( BO_i^1, BO_i^2, BO_i^3 \) and \( BO_i^2 \) denotes the backoff duration of \( v_i \) and \( v_j \) in frames \( F_1 \) and \( F_2 \) respectively. Then the probability that linking the new pseudonym of \( v_i \) with its old pseudonym by the MAC layer attack is equal to the probability that the messages transmitted by \( v_i \) do not mix with the other vehicles in the same frame. In this way, the messages of \( v_i \) would be taken as broadcasted in a virtual time slot that starts from the time instant \( v_i \) contends to access the channel. In this way, the lower bounded probability is calculated as:

\[
p_j = (1 - \frac{T_{BO}}{T_F})^{k-1}, \text{ where } T_F \text{ is the duration of the synchronized frame and } T_{BO} \text{ is the duration of the service time of the message. Based on the assumptions and analysis of the DSRC based BSM transmission in [16], when } k = 2, \text{ the probability is approximately to be } 1 \text{ and when } k = 60 \text{ in one-hop range, the data rate is 12Mps, packet size is 400 bytes and the contention window is 32, the probability } p_j \text{ is still nearly 0.5616 conditioned on the message is successfully transmitted, which is much larger than } \frac{1}{n}.\]

Consider an extreme case when \( k = 200 \), the probability would be lower bounded by 0.0614, which is still larger than \( \frac{1}{n} \). Therefore the linking uncertainty is much reduced by the MAC layer context information.

To deal with the context linking attack under the CSMA based MAC protocols, we propose that when vehicles generate the BSMs periodically in each frame, they distribute the messages transmission attempts during the frames uniformly. By this means, the transmission time of all the vehicles would be uniformly distributed in each frame, so that the linking uncertainty would be maximized to the adversaries. However, when exploiting the uniform transmission strategy in the CSMA based protocol, the packet delivery ratio is only about 60% when \( k = 60 \). Therefore, the transmission efficiency would be further improved by adding some coordination functions, e.g., coordinated time slotted based strategy, for the MAC layer protocol.

When vehicles exploit the coordinated time slotted based protocols, the attack is still possible to launch by collecting the slot utilization pattern. The adversaries can map the slot index with the current pseudonym of each vehicle, denote as \( M = \{ (s_1, PID_1), (s_2, PID_2), \ldots, (s_k, PID_k) \} \), where \( s_i \) and \( PID_i \) are the slot index and pseudonym of vehicle \( i \) respectively. If an adversary finds the pseudonym assigned for the transmissions in a specific slot has been changed, the adversary takes it as the alternation of the pseudonym by the current vehicle, rather than the entrance of a new vehicle. This is because when vehicle changing pseudonym, the time slot is kept continuously in transmitting. While if there is a new entered vehicle, the time slot will be idle for at least one broadcast period for the vehicle to notify itself to its neighbors in TDMA. Meanwhile, the vehicles in the same mix-zone need to apply different slots to avoid transmission collisions. Thus, the new map becomes \( \{ (s_1, PID'_1), (s_2, PID'_2), \ldots, (s_k, PID'_k) \} \). Obviously, if the slot utilisation and pseudonym changing operate inconsistently, the probability \( P_i \) becomes 1 to link \( PID_i \) with \( PID'_i \) via \( s_i \). Therefore the tracking uncertainty of the targeted vehicle \( v_i \) in the mix-zone becomes 0 and the adversary can track the vehicle based on the BSMs continuously.

The degree of location privacy not only depends on the location privacy achieved in mix-zones, but also on the maximum tracking time. Generally, the time-to-confusion is limited by the pseudonym age, if a targeted vehicle always changes the pseudonyms together with at least one neighbor in the mix-zone to confuse the adversary. However, when the MAC context linking attack is carried out, the time-to-confusion is much increased. Consider \( v_i \) reserves a time slot at the time \( t_0 \) and utilizes \( PID_i \) to transmit. It changes the pseudonym at time \( t_1 \) and keeps the occupancy of the slot until \( t_2 \). Therefore, \( \varepsilon_p = t_1 - t_0 \). However, when \( t_2 \) is not in the same tracking step as \( t_1 \) of the adversary, the time-to-confusion is increased to \( t_{conf} = \min \{ T_{trip}, T_{U>0} - t_0 \} \), where \( T_{trip} \) is the time to finish the trip and \( T_{U>0} \) is the time that the adversary becomes confused.

In other words, the location privacy of vehicles is achieved only if the following conditions are satisfied. 1) Each vehicle changes the pseudonym with enough cooperative vehicles in the mix-zones. 2) The MAC layer transmission pattern of each vehicle should be changed with the pseudonym simultaneously to resist the MAC layer context linking attack.

### 4 Distributed MAC Layer Attack Resistant Pseudonym (MARP) Scheme

MARP is a new cross-layer scheme which coordinates vehicles to adaptively change pseudonyms and access the
wireless channels in a distributed manner. The operation flow of each vehicle in MARP is illustrated in Fig. 3. It involves the time slots reservation to broadcast BSMs, mix-zone construction, pseudonym changing with new slot reservation. We assume that the system initialization, pseudonym provisioning, key generation, message signing, misbehaviour report and revocation are operated according to the standardized approach [14].

Before starts a trip, a vehicle can be preloaded a sufficient number of pseudonyms from the TA for the trip based on the estimation of the duration to the next TA on the trip. Each pseudonym has a lifetime that can be either fixed for all vehicles or flexible for each vehicle to achieve different privacy preferences.

### 4.1 Distributed Time Slotted Access

The proposed protocol is based on a slotted frame structure, as illustrated in Fig. 4. We design that every vehicle can occupy a dedicated slot in the distributed transmission (DT) period to broadcast its identity (including the pseudonym and certificate), safety information and the signature, and a list of its perceived status of all slots, which is referred to as the frame information (FI). Specifically, the kth time slot in the DT is denoted as \( s_k \) and its status information is denoted as \( s_{ki} \), which consists of two bits: “00” indicates the slot is perceived to be unoccupied; “01” indicates that the slot is occupied by a vehicle within the one-hop transmission coverage of the sensing vehicle (including the sensing vehicle itself); “10” indicates that the corresponding slot is occupied by a vehicle outside the one-hop transmission coverage of the designated vehicle but within that of a one-hop neighbor of the sensing vehicle, i.e., a hidden node to the sensing vehicle; and “11” indicates a transmission collision, i.e., multiple concurrent transmissions are received at the vehicle. Moreover, a contiguous number of slots towards the end of a frame can be released from being occupied by specific vehicles, and can be combined for contention-based access, e.g., by exploiting DIFS.

In addition to the periodic broadcast of safety messages, vehicles dynamically transmit the encrypted pseudonym requests to the roadside TAs based on their locations and the residual trip duration. Vehicles may also need to receive the CRL and report misbehaviour by communicating with the roadside TAs. To separate the transmission of these occasional messages to the periodic safety messages, a contention based flexible schedule (FS) period. As a consequence, the pseudonym request timing information does not harm the user privacy and the occasional transmission does not harm the safety messages transmission. To this end, the frame structure is of strong compliance with a standard IEEE 802.11 frame which can accommodate two separate periods of point coordination function (PCF) and distributed coordination function (DCF).

When \( v_i \) starts to acquire a new slot or shuffle the old slots, it initiates a random access procedure to reserve a dedicated slot that has not been occupied by all its two-hop range neighbours, and hence the hidden terminal issue is prevented. To achieve this, vehicle \( v_i \) needs to carry out a procedure to update the FI by listening to the channel for a complete transmission period to receive the messages from its one-hop neighbors, extract their FI, and identify unoccupied slots within the transmission range of its one-hop neighbors and their one-hop neighbors (which can contain the hidden nodes to \( v_i \)). After updated the unoccupied slot set, vehicle \( v_i \) randomly selects an available time slot indexed \( t_{si} \) and transmits its own messages as illustrated in the bottom of Fig. 4. The format of the message generated by the tagged vehicle \( v_i \) is \( m_{si} = \langle \text{Header}, |F|, \text{Cert}_{P\text{ID}_i}, k_{ij}(M) \rangle \), where \( M \) is the safety payload, \( \text{Cert}_{P\text{ID}_i} \) is the pseudonym certificate of vehicle \( v_i \) corresponding to the jth pseudonym and the signature \( \text{Sign}_{k_{ij}}(M) \) is the signature of message \( M \) generated using the private key \( k_{ij} \) based on the Elliptic Curve Digital Signature Algorithm (ECDSA).

The vehicle \( v_i \) can be notified if the slot is successfully reserved by listening to the channel for the next \( N_{ts} - 1 \) transmission slots. Specifically, if all the one-hop neighbours of \( v_i \) have updated that status of \( t_{si} \) is “01”, the reservation of the slot dedicated to \( v_i \) is successful. There is a possibility that there is at least another vehicle, for example, \( v_j \) (\( j \neq i \)), attempts to access the same slot as \( v_i \). As a consequence, a transmission collision occurs. The vehicles \( v_i \) and \( v_j \) can be either within the one-hop transmission range of each other, or hidden nodes towards each other. In the case that they are within the one-hop transmission range of each other, the collision can be clearly indicated in the slot occupancy information as “11” (collide) from the following transmissions of other vehicles. In the case that \( v_i \) and \( v_j \) are hidden nodes to each other, some of their one-hop neighbors may update the slot occupancy to indicate that the slot is successfully reserved (by either of these two vehicles), while others indicate a collision in the slot. In both of the cases, the “collide” indicator of the slot can be perceived. The reservation of the slot is unsuccessful for
both vehicles. The vehicles restart their reservation process again, and repeat until they successfully reserve the slots. At the end of each time slot, each vehicle should update its neighbour set and the status of the time slots, and broadcast the updated FI in its own messages. The pseudocode of the time slot reservation is described by the Algorithm 1.

Algorithm 1 Time Slot Reservation

1: for each vehicle $v_i$:
2: during each slot $t_{sk}$ in the DT:
3: sets $s_k$ as “00”
4: if received exactly one message then
5: sets $s_k$ as “01”
6: else if received from a one-hop neighbour saying $s_k$ is “01” then
7: sets $s_k$ as “10”
8: else if received more than one message from different vehicles or a collision then
9: sets $s_k$ as “11”
10: end if
11: updates the FI and randomly selects a new transmission slot $t_{si}$ with status “00” during the next frame
12: broadcasts the BSM in $t_{si}$
13: waits for $N_{ts} - 1$ slots following $t_{si}$
14: if all the neighbours update $s_i$ as “01” then
15: considers the reservation is successful
16: else
17: does the algorithm until it reserves a transmission slot successfully
18: end if

It is also possible that vehicles, previously beyond the two-hop coverage of each other and hence occupied the same slot, move towards each other, become hidden nodes, and cause transmission collisions (referred to as “merging collision”) at the slot. By monitoring the lists of slot occupancy from surrounding vehicles, the vehicles of interest can become aware of the collisions, release the occupancy of the slot, and start the reservation process again, as described above. Each vehicle will utilize the slot in each frame to broadcast the safety messages periodically and release the time slot until it changes pseudonym or detects a merging collision. The pseudocode for the time slot release procedure is illustrated in Algorithm 2.

Algorithm 2 Time Slot Shuffle

1: for each vehicle $v_i$:
2: updates the slot status list after each time slot,
3: if (the $s_i$ is updated by all the neighbours as “01”) & (does not need to construct a mix-zone) then
4: broadcasts the BSMs periodically in the current time slot
5: else
6: releases the current slot and does Algorithm 1 until it reserves a new transmission slot in the DT
7: end if

As discussed, the decentralized slotted access is able to eliminate the hidden issue, and embrace the mobility of vehicles. Moreover, it is able to guarantee collision-free delivery of critical safety information in a timely fashion, which is of practical value to many mission critical applications and scenarios, such as columns of military vehicles in war zones.

4.2 Mix-zone Construction

Both the reserved slots and pseudonyms of the vehicles need to be changed regularly. This is because they can be correlated and associated with particular vehicles, and the use of them for an extensive period of time would provide adversaries opportunities to identify each individual vehicle and track its trajectory. It is particularly important to change the slots and pseudonyms of multiple vehicles together to confuse an adversary. For this reason, we assume that every pseudonym preloaded at a vehicle is assigned with a predefined lifetime for the effective use of the pseudonym.

We design a fully distributed strategy for that each vehicle makes the decision independently. Each vehicle listens to the safety messages broadcasting by its neighbors and determines whether to construct the mix-zone based on its own pseudonym age and the cooperation of its neighbors. Before each transmission, each vehicle checks the number of vehicles within its one hop set and whether its pseudonym age is equal to or larger than the cooperation age threshold, i.e. $\varepsilon_p \geq \tau$. The vehicles can set different cooperation age thresholds, so as to achieve their preferred location privacy levels. If it is, $v_i$ enters a semi-silent period by transmitting the BSMs with null location information and sets the cooperation indicator $\text{ChangeReq}$ continuously until it meets at least one cooperative vehicle. By this way, a consensus is implicitly made among the cooperative vehicles to construct the mix-zone. After the mix-zone is constructed, the cooperative vehicles randomly select a silent period from $[t_{s_{min}}, t_{s_{max}}]$ to keep silent. During the silent period, there will be more vehicles driving into the mix-zone. As a consequence, the anonymity set size and the available slot set size are both increased. After the silent period expires, the vehicle performs the slot reservation process using a new pseudonym. The pseudocode for the mix-zone construction procedure is shown in Algorithm 3.

Algorithm 3 Mix-zone Construction

1: for each vehicle $v_i$:
2: updates the age of the current pseudonym
3: if the pseudonym age $\geq \tau$ then
4: transmits the BSM with null safety information and the $\text{ChangeReq}$ indicator
5: if $|C_v| \geq 1$ then
6: randomly chooses a silent period within $[t_{s_{min}}, t_{s_{max}}]$
7: released the current transmission slot
8: remains silent until the silent period expires
9: uses a new pseudonym to implement Algorithm 1 until it reserves a new time slot
10: end if
11: end if
preferences.

Firstly, the RSUs can cooperate with the vehicles to adjust the duration of the transmission period and the schedule period, so as to adjust the frequency of the broadcast of the safety messages and the security related messages, i.e., new pseudonyms, CRLs and misbehaviour reports.

Secondly, even without the RSUs, vehicles can detect the traffic density accurately by updating the status of the transmission slots. When the traffic is congested, decentralized congestion control schemes are easily to apply, e.g., vehicles can decrease the transmission frequency by taking turns with their neighbours in the same time slots to broadcast. On the other side, when the vehicles detect the traffic density is too sparse, vehicles can choose a longer silent period when construct the mix-zone to gather the cooperative neighbours, so as to protect the location privacy with moderate sacrifice on the traffic safety.

In other words, the propose protocol is adaptable to different traffic scenarios and support different levels of locatio privacy and transmission requirements. While in this paper, we consider the free-flow traffic scenarios and the cooperation age threshold of all vehicles is assumed to be the same.

5 Analytical Evaluation

5.1 Performance Metrics

In this section, we analytically evaluate the location privacy quality of MARP. The metrics include the pseudonym age, average anonymity set size and time-to-confusion. In addition to the three privacy metrics, the packet overhead and the frame duration are also taken into consideration to evaluate the efficiency of the scheme.

- The pseudonym age $\varepsilon_p$ is defined as the interval during which a pseudonym is used.
- The probability $p(c)$ that the anonymity set size is larger than 1 in a mix-zone is derived, which is the probability that at least two vehicles cooperate to construct a mix-zone. Meanwhile the average anonymity set size of the mix-zone is calculated.
- The time-to-confusion $t_c$ is defined as the successive duration that an adversary can distinguish and trace a target vehicle. It depends on the pseudonym age, the anonymity set size, the silent duration before changing the pseudonym and the safety message transmission rate of vehicles in the mix-zone.

5.2 The Age Fluid Model

We propose a novel analytical model named as the age fluid model to describe the evolution of the pseudonym age of the vehicles in the system. By mean field theory, we note that, as the number of vehicles becomes large, although their individual pseudonym age is time-varying, the pseudonym age distribution of the whole system is asymptotically deterministic [34].

Define the pseudonym age set of the system at time $t$ is $\varepsilon_p(t) = (\varepsilon_p(i: i=1)^N)$, whereas $\varepsilon_p(t)$ characterizes the most recent pseudonym age of vehicle $v_i$, $N$ is the total number of vehicles within the same geographical area. The pseudonym age for a vehicle $v_i$ is calculated as $\varepsilon_p(t) = t - t'$, where $t$ is the current time, and $t'$ is the time of the last pseudonym change. The dynamics of age is characterized by the aging and jump process [26]. When at least two vehicles decide to construct the mix-zone together, they change the pseudonyms and the age jumps to 0. The rate of the jump process depends on the number of vehicles within the transmission range, and the cooperation probability of the vehicles. Otherwise, the pseudonym age of a vehicle experience the aging process, in which it increases in a rate of the BSM broadcasting. The pseudonym aging in the drift process depends on the transmission interval of the BSMs transmission. Therefore, the higher transmission frequency, the larger aging rate, to characterize the effect of the BSMs to the location privacy. If $v_i$ decides to change the pseudonym age at time $t$, it waits for the cooperative neighbours. Denote the set of cooperative neighbours as $C_i$. $\gamma$ is the time cost for the pseudonym changing, therefore the new pseudonym starts its lifetime from $t + \gamma$.

$$\begin{cases}
\varepsilon_p(t + \gamma) = 0, t' = t, \text{ when } |C_i| > 0 \\
\varepsilon_p(t) = \varepsilon_p(t' - \gamma), \text{ when } |C_i| = 0
\end{cases} \tag{2}$$

The pseudonym changing cost $\gamma$ of $v_i$ is defined as the sum of the semi-silent period to hear from at least one cooperative neighbor and the silent period. The cost is expressed in the aging rate which models the gap between the jump process and a new round aging process of the pseudonyms of each vehicle. As the pseudonym age distribution of each vehicle is tight, the evolving of the whole system becomes deterministic when the number of vehicles is large in the system. We define the occupancy measure of vehicles in the network with the pseudonym age of $z$ by

$$M(z, t) = \frac{1}{N} \sum_{i=1}^{N} \delta_{\varepsilon_i(z, t)}.$$  \tag{4}$$

$M(z, t)$ also denotes the density of vehicles with the pseudonym age of $z$, i.e. $f(z, t)$. Therefore, the cumulative distribution function (CDF) over $M(z, t)$ is:

$$F(z, t) = M(t)([0 : z]) = \int_{0}^{z} M(\varepsilon_p, t) d\varepsilon_p \tag{3}$$

$F(z, t)$ denotes the proportion of $N$ with pseudonym age less than or equal to $z$. Based on the mean field limits, when $N$ is large enough and the mobility of the vehicles is independent from each other, the evolving-stationary model indicates that the collection of the occupancy measure of the pseudonym aging of vehicles $M(z, t)$ converges in distribution of deterministic process $\{\tilde{m}(t) | t \geq 0 \}$. Therefore, the evolving flow of the pseudonyms can reach a stationary rate, i.e. $\frac{\partial F(z, t)}{\partial t} = 0, t \rightarrow \infty$. Here we will work out the distribution of $\{\tilde{m}(t)\}$ by characterizing the dynamical changes of the occupancy measure for the age of pseudonyms less than or equal to $z(z > 0)$ in the network in a small interval $dt$.

A fraction of pseudonyms experience the aging process and their age grows older in the interval $dt$. However, only the fraction of vehicles’ pseudonyms with $z - \Delta z$ grow older and become older than $z$, and need to be removed from $F(z, t)$. While the others do not need to be subtracted, as they are not in $F(z, t)$ before the interval $dt$. Therefore, the rate of change of $F(z, t)$ caused by the aging process is calculated as:

\[ \frac{\partial F(z, t)}{\partial t} = \frac{\partial F(z, t)}{\partial z} \frac{\partial z}{\partial t} = \frac{\partial F(z, t)}{\partial z} \frac{\partial z}{\partial \varepsilon_p} \]
\[
\lim_{\Delta z \to 0} \frac{F(z - \Delta z, t) - F(z, t)}{\Delta z} = \frac{\partial F(z, t)}{\partial z} = M(z, t)
\] (4)

Meanwhile, the rest vehicles' pseudonyms experience the jump process, in which the vehicles meet with each other and decide to change the pseudonyms in the mix-zone. If the vehicle's old pseudonym age at time \( t \) is less than \( z \), then after the pseudonym changing within \( \partial t \), the pseudonym age is reset to 0, which is still less than \( z \). Therefore, it causes no change of the \( F(z, t) \). If the vehicle's old pseudonym age at time \( t \) is equal to or larger than \( z \), and it decides to change the pseudonym age within \( \partial t \), the pseudonym age is reset to 0, which causes an increase of \( F(z, t) \). Consider the time cost by the semi-silent period to gather the cooperative neighbours is upper bounded by \( \gamma_s \), the changing rate of \( F(z, t) \) caused by the jump process is calculated as:

\[
\int_z^\infty p(c(t))f(\varepsilon_p, t)d\varepsilon_p = \int_{\tau}^{\tau + \gamma_s} p(c(t))f(\varepsilon_p, t)d\varepsilon_p
\] (5)

where \( p(c(t)) \) is the probability that at least one of the encountered vehicles chooses to cooperate. It can be calculated as the combination of the probability of having \( n \) vehicles in the surrounding area, and the probability that at least one of them cooperates:

\[
p(c(t)) = 1 - \sum_{n=0}^{\infty} p_n(1 - c(t))^n
\] (6)

where \( p_n \) is the probability of having \( n \) neighbors within the one-hop set of a vehicle \( v_i \). \( c(t) \) is the cooperation probability of a neighbor. Assume the network is homogenous that all the vehicles apply the same age threshold, therefore the cooperation probability is calculated as:

\[
c(t) = \int_0^\infty c(z)f(z, t)dz = \int_{\tau}^{\tau + \gamma_s} f(z, t)dz
\] (7)

\[
c(z) = \begin{cases} 0, & z < \tau \\ 1, & z \geq \tau \end{cases}
\] (8)

Consequently, for the evolving stationary system, unique solution of the differential equation \( \frac{\partial F}{\partial t} \) here is:

\[
\left\{ \begin{array}{l}
\frac{\partial F}{\partial t} = -\frac{\partial F(z, t)}{\partial z} + \int_{\tau}^{\infty} p(c)f(\varepsilon_p, t)d\varepsilon_p \\
F(\infty, t) = 1, \forall t \geq 0
\end{array} \right.
\] (9)

To solve Equation (8) and calculate the age distribution of pseudonyms, the stable boundary condition \( \frac{\partial F}{\partial z} = 0, t \to \infty \) is considered. Consequently, the equation can be calculated based on the above stationary regime, Equation (4-8) and \( \frac{\partial F}{\partial z}(z, t) = f(z, t) \). Consequently, we obtain,

\[
\left\{ \begin{array}{l}
\frac{\partial f}{\partial z} + p(c)f(z)c(z) = 0 \\
\int_0^\infty f(z)dz = 1
\end{array} \right.
\] (10)

when \( z < \tau \), the Equation (9) becomes \( \frac{\partial f}{\partial z} = 0 \). Therefore \( f(z) = f(0) \). When \( z \geq \tau \), the equation is \( \frac{\partial f}{\partial z} + p(c)f(z) = 0 \). Considering the boundary \( z = \tau \), we obtain \( f(z) = f(0)e^{-p(c)(z-\tau)} \). And the final solution becomes:

\[
f(z) = \begin{cases}
\frac{1}{\tau p(c)}, & 0 \leq z < \tau \\
\frac{e^{-p(c)(z-\tau)}}{\tau p(c)}, & z \geq \tau
\end{cases}
\] (11)

Therefore, by adjusting the value of \( \tau \), different privacy levels can be achieved in the network. Based on the cooperation function of \( c(z) \) and \( f(z, t) \), \( c(t) \) is calculated as:

\[
c(t) = \frac{1}{1 + \tau p(c)}.
\] (12)

Thus, if the cooperation probability \( p(c) \) is calculated, the final result of \( f(z, t) \) can be achieved. Meanwhile, the average anonymity set size of a mix-zone can be calculated.

### 5.3 Derivation of the Cooperation Probability

To calculate \( p(c) \), the probability \( p_n \) needs to be determined. Consider the traffic is under a balanced steady flow condition that the arrival rate is \( \lambda \) that follows the Poisson distribution. The transmission range \( R \) of each vehicle is assumed to be much larger than the width of the roads. According to [34], the speed per vehicle is identically and independently distributed yielding a truncated normal distribution. The probability density function (pdf) is given as

\[
g(s) = \frac{\xi}{\sigma \sqrt{2\pi}} e^{-\frac{(s-\bar{s})^2}{2\sigma^2}}
\] (13)

where \( \bar{s}, s_{min}, s_{max} \) and \( \sigma \) are the average, minimum, maximum and the standard deviation of vehicle speeds respectively. Based on the well established traffic flow theory that, over a specific road segment, the average traffic density is given by

\[
\rho = \lambda \bar{s}^{-1}
\] (14)

Given vehicles distributes according to the spatial Poisson process as: \( Pr(N = n) = \frac{(2R\rho)^n}{n!} e^{-2R\rho} \), with the expected number \( \lambda_{2R} = 2R\rho \). The probability of \( v_i \) having \( n \) neighbors is equal to the probability that the total number of vehicles within the specific area is \( N = n + 1 \). Thus the probability mass function (pmf) of \( |N(v_i)| \) is given as:

\[
Pr(|C_{v_i}| = n) = Pr(N = n + 1) = \frac{\lambda_{2R}^n e^{-\lambda_{2R}}}{(n+1)!}, n = 1, 2, ...
\] (15)

Consequently, Equation (5) is calculated:

\[
p(c) = 1 - \sum_{n=0}^{\infty} \frac{\lambda_{2R}^n e^{-\lambda_{2R}}}{(n+1)!} - 1 = 1 - \sum_{n=0}^{\infty} \frac{\lambda_{2R}^n e^{-\lambda_{2R}}}{n!} (1 - c(t))^n
\] (16)

By replacing \( c(t) = \frac{1}{1 + \tau p(c)} \) by \( p(c) \) in the Equation (14), \( p(c) \) is the value between 0 and 1 that satisfies the (14). Thus it can be solved by the iterative methods.

\[
p(c) = 1 - \frac{1 + \tau p(c)}{\tau p(c)} e^{-\lambda_{2R} \frac{1}{\tau + \tau p(c)}} - e^{-\lambda_{2R}}
\] (17)
Each vehicle maintains a constant speed and moves with negligible interaction with its neighbors within an observation period. Thus it is practical to assume that the one-hop neighbour set of \( v_i \) stays constant when construction a mix-zone, and can be calculated based on vehicle density distribution in the network. If the number of the one-hop neighbours is known, correspondingly, Equation (15) is approximately calculated as \( p(c) = 1 - (1 - \frac{\gamma}{1 + \tau p(c)})^n, n \geq 1. \)

### 5.4 Final results of the privacy metrics

Based on the derivation of the distribution of the pseudonym age, we can calculate the average pseudonym age based on the assumption that the system is stationary and uniform. The upper bound of the pseudonym age can be calculated out from Equations (10) and (15) as

\[
\bar{\varepsilon}_p = \min\{z | f(z) = 0, z \geq \tau\} \tag{17}
\]

According to the evaluation, the average anonymity set size within a specific area depends on the vehicle cooperation rate, the traffic density on the roads, the pseudonym age threshold and the silent period \( ts \). The average target neighbour set size is: \( \frac{N_v}{1-c^{-\frac{1}{\gamma}}}. \) As the cooperation decision of each vehicle is independent to each other, the average anonymity set size is calculated as: \( E[C(v_i)] = \frac{N_v}{1-c} \cdot \min(1, \frac{\gamma}{\epsilon} + \frac{\gamma_{\text{max}}}{2}). \)

As the silent period duration is uniformly distributed, the expected silent period of each vehicle is \( \frac{\epsilon_{\text{min}} + \epsilon_{\text{max}}}{2} \). Based on the analysis in the subsection 3.3, the traceability of a targeted vehicle \( v_i \) will become zero when the target vehicle changes the pseudonym and releases the transmission slots simultaneously. In this way, the time-to-confusion can be bounded by the pseudonym age.

\[
t_c = \min\{\bar{\varepsilon}_p, t_{\text{trip}}\} \tag{18}
\]

In contrast, if the time-to-confusion of an adversary is not bounded by the pseudonym age in a pseudonym scheme, the pseudonym scheme is taken to fail to guarantee the location privacy. As the influential factors include the cooperation age threshold, traffic density and the silent period, more effective pseudonym schemes can be designed based on the framework of the MARP scheme and the analytical model by designing the cooperation age threshold and silent period for vehicles to achieve the preferred bounded time-to-confusion.

### 5.5 Packet Overhead Evaluation

As described in Section 4, each safety message transmitted in a slot consists of the header, BSM payload, certificate with the corresponding pseudonym and the signature. The format of the safety message is shown in Fig.3. The size of the BSM is calculated as follows.

\[
L_m = B_{N_v} + B_{\text{header}} + B_{\text{payload}} + B_{\text{sign}} + B_{\text{cert}} \tag{19}
\]

where \( B_{N_v} \) is the number of bytes to denote the status of each slot respectively in the FI, \( B_{\text{header}} \) is the number of bytes of the header specified for the WAVE safety message, \( B_{\text{payload}} \) is the number of bytes of the safety data, \( B_{\text{sign}} \) is the number of bytes required for the signature, \( B_{\text{cert}} \) is the number of bytes required for the certificate. Assume \( N_v = 200 \), corresponding to the most congested scenario in which the number of vehicles in one hop set is 80, therefore the two hop slot usage ratio is 0.8. It requires 50 bytes to represent the slot status. \( B_{\text{sign}} = 56, B_{\text{header}} = 19, B_{\text{payload}} = 53, B_{\text{cert}} = 126. \) Therefore, the total length a BSM is 304 bytes. The delay that a safety message experiences on the channel depends on the duration of a transmission slot. To evaluate the appropriate interval of the slot, considering a mandatory supported transmission rate 12 Mbps by IEEE 802.11p, the transmission delay takes less than 0.22 ms, the transmission period is assumed to be 40 ms and the flexible schedule period is assumed to be 6 ms, which is assumed to be long enough to support the occasional transmission. The synchronized frame duration is compliant with the multi-channel operation with the CCH interval as 46ms.

### 6 Analytical Evaluation and Simulation Results

In this section, we firstly verify the accuracy of the analytical model, and then we carry out extensive simulations using a city scenario to validate the performance of the MARP scheme by comparing with other pseudonym schemes.

#### 6.1 Analytical Analysis and Model Validation

In this subsection, we evaluate the analytical results presented in the previous subsection. In order to verify the accuracy of the analytical model, we presents the Matlab and C++ simulation results compared with the numerical results.

We consider a 2-km highway scenario with two opposite directions as described in the analytical model. The average traffic density is assumed to be from 0.0125 to 0.1 with the increment step as 0.0125 vehicles per meter per lane. Each vehicles moves with a constant speed drawn from the normal distribution and the number of vehicles on the system is kept constant during the simulation. To analyze the performance of the stable system, the vehicles will start to move when they reserve the transmit slots successfully. When a vehicle reaches one end of the road, it reenters the same lane from the entrance point, and starts to increase the age of its current pseudonym when transmitting the new safety message. Assume the physical channel is ideal so that each vehicle can communicate with all the vehicles within its transmission range with no obstacles. The transmission range of each vehicle is assumed to be 200m, therefore the expected number of vehicles within the one-hop range is from 10 to 80. The other simulation parameters as summarized in Table 2.

First, the probability \( p(c) \) of at least one vehicle cooperating in one frame interval is observed in Fig. 5. As we can see that the cooperation probability is rather low in a synchronized broadcast interval. The reason is that the pseudonym age of vehicles is approximately uniformly distributed, so that if one vehicle decides to change the pseudonym in one frame, it needs to wait and spends a silent period to gather more neighbours with their pseudonym age evolved for the cooperation. We evaluate the influence of pseudonym...
age threshold and the traffic density. As the number of neighbours in the targeted area (i.e. one hop transmission range) increases, the cooperation probability $p(c)$ increases logarithmically. When the pseudonym age threshold increases, we observe that $p(c)$ decreases for any value of $\tau$. The reason is that for larger value of $\tau$, a larger fraction of vehicles will have an age of pseudonym below $\tau$ at any constant. For this reason, for a high age threshold, fewer vehicles cooperate at a specific instant, and consequently $p(c)$ decreases.

Fig. 6 compares the average pseudonym age (Fig. 6 (a)) and the average anonymity set size (Fig. 6 (b)) calculated in the analytical model (denoted as Ana) to the results obtained by simulations (denoted as Sim) with five different cooperation age thresholds. The value of the average pseudonym age obtained from the analytical model presents a pretty accurate match with the simulation results. Although the cooperation probability is rather low in a frame interval, the silent period in the mix-zone facilitate the increasing of the anonymity set size. As $\tau$ increases, we observe that the average pseudonym age increases accordingly. The pseudonym age is mainly dominated by the cooperation pseudonym threshold, as its value is close to the threshold with no larger than 10 seconds. As the number of vehicles in the targeted area increases, the average pseudonym age decreases slightly as more vehicles are cooperative, making the jump process in the system occur more frequently. Fig. 6 (b) shows the anonymity set size in the mix-zones obtained in MARP. The effectiveness of MARP is validated even under sparse traffic density scenarios. The anonymity set size decreases with the increase of $\tau$. The results are rational and obvious, as with a large $\tau$, $p(c)$ decreases and fewer vehicles choose to cooperate. In addition, a dense traffic has a positive impact on the anonymity set size. It is also the reason that some researchers propose to change pseudonyms when vehicles detecting traffic congestion or driving into large intersections [3]. The simulation results of the anonymity set size are a little higher than the analytical results. The reason is that during the simulation, the number of vehicles is limited so that their pseudonym age is not independently distributed and becomes similar after several rounds of mix-zone construction.

### 6.2 City Scenario and Simulation Protocols

In addition to the verification of the MARP protocol based on the same assumption of the analytical model, we further study the performance of MARP using a city grid layout scenario. It consists of four square city blocks and intersections of a horizontal street with a vertical one. Each street has two directions and vehicles with larger speed can go ahead of the vehicles with the lower speed. We assume when a vehicle reaches a junction area, it chooses on of all possible directions randomly. These vehicles always move within the simulation area and never drive out. The intersection range of each side is assumed to be 50m. The total number of vehicles is from 120 to 960 with the step as 120, corresponding to average 10 to 80 vehicles within one hop set.

At the beginning of the simulations, vehicles are uniformly placed on the streets. Vehicles remain stationary and randomly select a time instant within a interval as long as the cooperation age threshold to acquire a time slot by using the designed protocol. Once the network reaches a steady state, the vehicles start to move and the simulation timer begins. The pseudonym age starts to count from the vehicle transmits the first safety message. We measure the average pseudonym age, anonymity set size and the time-to-confusion of MARP under different traffic densities and two different pseudonym cooperation thresholds, $\tau = 2$, 5 minutes as the recommended pseudonym age of the SAE J2735 [15] and European standard [2] respectively.

Considering the cross-layer performance, the packet delivery ratio (PDR) is used to verify the transmission reliability for MARP. The PDR of a vehicle is calculated as the total number of BSM messages that have been successfully transmitted within the lifetime to the total number of the BSMs generated by the vehicle. During a vehicle releases the old transmission slot and keeps silent before the new transmission slot reservation, the messages are assumed to be undelivered. When a vehicle encounters merging collisions or does not reserve a time slot, the safety messages are also assumed to be undelivered. The transmission delay is an important metric in the measurement of transmission efficiency. Using the TDMA based MAC layer operation, safety messages are always broadcasted in the specific transmission slots, thus the transmission delay in the MAC layer

#### TABLE 2

Parameter Settings in Numerical Evaluation Verification

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean vehicle speed</td>
<td>40km/h</td>
</tr>
<tr>
<td>Vehicle speed standard deviation</td>
<td>10km/h</td>
</tr>
<tr>
<td>The length of the street</td>
<td>800m</td>
</tr>
<tr>
<td>The width of the street</td>
<td>5m</td>
</tr>
<tr>
<td>$\tau$: cooperation pseudonym age threshold</td>
<td>[1, 5] minutes</td>
</tr>
<tr>
<td>Safety message broadcast frequency</td>
<td>10Hz</td>
</tr>
<tr>
<td>$t_{\text{min}}$: the minimum silent period</td>
<td>2.5 seconds</td>
</tr>
<tr>
<td>$t_{\text{max}}$: the maximum silent period</td>
<td>7.5 seconds</td>
</tr>
<tr>
<td>$t_{\text{rep}}$: Driving duration</td>
<td>20 minutes</td>
</tr>
<tr>
<td>The number of lanes on the road</td>
<td>2</td>
</tr>
</tbody>
</table>

#### Fig. 5. Influence of vehicle number and pseudonym age threshold: Probability of at least one neighbor cooperates

![Cooperation Probability vs Total Number of Vehicles within One Hop Set](image)
is bounded within 100ms.

The performance of the compared schemes without the collaborative MAC layer attack resistance strategy is also measured in this paper. In the compared schemes, vehicles using the same MAC operation to transmit as in MARP. We consider two pseudonym changing schemes. In the first scheme N-MARP, vehicles decide to construct the mix-zones based on their pseudonym age distributively as in MARP. However, they just change the pseudonyms after the silent period, but do not shuffle the transmission slots. Therefore, the number of pseudonyms consumed by each vehicle in the MARP and N-MARP schemes during the trip is based on the pseudonym lifetime. While, the second scheme applies the effective pseudonym changing at social spots (PCS) strategy [3]. In PCS, vehicles stop transmitting the safety messages when they enter the intersection area, and begin to construct the mix-zones. After they leave the intersections, they start to utilize the new pseudonyms to transmit safety messages.

6.3 Simulation Results

The average anonymity set size in the mix-zones achieved in the three schemes is presented in Fig. 7. It shows that the PCS strategy achieves the largest anonymity set size. In PCS, all the vehicles gathering in the intersections change the pseudonyms together without the consideration of the pseudonym age. The total number of vehicles at road intersections is much larger than the number of vehicles with the pseudonym lifetime expired within the one hop range in the MARP and N-MARP schemes. All the three schemes can achieve a guaranteed anonymity set size level in the mix-zones. As the pseudonym changing strategies of MARP and N-MARP are both based on the cooperation age threshold, they show almost the same performance of the anonymity set size. In the figure, it is demonstrated that the anonymity set size decreases in MARP and N-MARP as the increase of the cooperation pseudonym age threshold, which also correlates with the analytical results.

In Fig. 8, the pseudonym age in the three schemes are investigated. As the figure shows, the pseudonym age in MARP and N-MARP is mainly determined by the cooperation age threshold. In contrast, the pseudonym age in PCS is influenced by the frequency of encountering intersections. In the simulated scenario, the average traveling duration between two intersections is about 1 minute. Therefore, vehicles consumes more pseudonyms in PCS in the simulated scenario. While in the real traffic environments, the utilization of pseudonym relies on the traffic topologies and traveling routes. If vehicles encounter intersection too frequently, it causes a waste of pseudonyms. On the other hand, if the intersection distribution is too sparse, it causes a long tracking duration of vehicles by adversaries and even compromises the location privacy.

As Fig. 9 illustrates, the maximal time-to-confusion is unbounded by the pseudonym age in N-MARP and PCS
regardless of the pseudonym changing. In N-MARP, vehicles change pseudonyms without shuffling the transmission slots. Vehicles usually acquire separately slots within two-hop range, and the mix-zone construction of each vehicle is carried out with its one hop neighbours. Thus, the adversary can distinguish each vehicle based on its transmission slot index. Consequently, the time-to-confusion is much larger in N-MARP and PCS than in the MARP protocol. The time-to-confusion in PCS shows a slight decrease when compared with N-MARP. The reason is that in PCS, vehicles change pseudonyms at intersections, where the merging collisions happen more frequently. Thus, there is a higher probability that the pseudonym changing and slot releasing happen simultaneously among several vehicles in PCS, causing the adversaries lose the linking of the pseudonyms. However, the adversaries are still able to track a targeted vehicle continuously at most cases in PCS. In the proposed MARP protocol, vehicles change the pseudonyms with the transmission slots during the silent period, which cuts the linking of the new pseudonym and old pseudonyms of vehicles in the network.

Fig. 10 compares the average packet delivery ratio of the vehicles during the their trip. Both the collisions and the silent period spent during the mix-zone construction degrade the performance of the safety message delivery. As in N-MARP and PCS, vehicles change pseudonyms without shuffling the transmission slots. The transmission slots can be utilized until merging collisions happen. Therefore, the performance of the packet delivery ratio is only degraded by the silent period and the merging collisions caused by the traffic mobility. Meanwhile, the PCS scheme exploits the minimum silent period at the intersections, where vehicles still send the messages but set the location and speed data to 0 [3]. Therefore, PCS has the highest packet delivery ratio. In the proposed MARP protocol, the packet delivery ratio is influenced by the pseudonym changing. When vehicles cooperate to change pseudonyms, they need to release the transmission slots and reserve the transmission slots again after the silent period. Therefore, the proposed MARP protocol has the worst packet delivery ratio when vehicles change pseudonyms more frequently ($\tau = 2$ minutes). As the number of vehicles increases, the merging collisions dominate the packet loss in the network, as a consequence, the final packet delivery ratio trends in the simulation results indicate that the packet delivery ratio decreases nearly in an exponential trend as the number of vehicles increases. Although the proposed scheme sacrifices the traffic safety for the location privacy, the achieved packet delivery ratio is till nearly 80% in the extreme high traffic density when the cooperation threshold is 2 minutes. Therefore, the protocol presents an effective performance in both the location privacy and safety messages transmission.

7 Conclusion

In this paper, we have introduced a new MAC layer context linking attack that may compromise the location privacy in VANETs and pointed out that the pseudonym changing schemes must be designed in a collaborative manner with the MAC layer protocols. To deal with the new attack, a new cross-layer scheme, named MARP, is proposed in this paper. In the MARP scheme, vehicles adaptively change pseudonyms and access the wireless channels in a distributed manner. In particular, we have developed an analytical model to formally analyse the achieved location privacy in the MARP scheme, in terms of the pseudonym age, anonymity set size and time-to-confusion. The analytical model is general to be applied for both CSMA and TDMA based network and can be utilized to analyze other pseudonym schemes. The analytical evaluation and simulation results verify the effectiveness of the scheme. To the best of our knowledge, most previous work considered the MAC layer protocol and pseudonym schemes separately. This paper shed light on the research of cross-layer location privacy protection in VANETs. Our future work is to explore more specific privacy preservation techniques under the framework of MARP by considering various traffic scenarios and location privacy preferences of different users.
Fig. 10. The average PDR of MARP, N-MARP and PCS schemes

References


[2] ETSI TS 102 867 v1.1.1-intelligent transport systems(ITS); security; stage 3 mapping for ieee 1609.2[J]. Standard, TC ITS, June 2012.


Zishan Liu [S’18] received her B.S. degree from the International College, Beijing University of Posts and Telecommunications (BUPT) in 2012. Currently she is working toward Ph.D. degrees at both BUPT and Macquarie University. Her research topics include vehicular communications, 5G wireless networks, network security and privacy preserving.

Zhenyu Liu [S’18] received his B.S. degree from the Henan University (HENU) in 2014. After that, he joined the School of Information and Communication Engineering in Beijing University of Posts and Telecommunications (BUPT) to pursue the Ph.D. degree. His research interests are in the area of security protection in vehicular networks and physical layer security.

Lin Zhang [M’09] received the B.S. and Ph.D. degrees from Beijing University of Posts and Telecommunications (BUPT), China, in 1996 and 2001, respectively. He was a Post-Doctoral Researcher with Information and Communications University, South Korea, from 2000 to 2002. He held a research fellow position with Nanyang Technological University, Singapore, from 2003 to 2004. He joined BUPT as a Lecturer in 2004, and then became an Associate Professor in 2005 and a Professor in 2011. He is currently serving as the Dean of the School of Information and Communication Engineering, BUPT. His current research interests include resource management in interconnected wireless/wired networks, mobile cloud computing, vehicular ad hoc and sensor networks, and Internet of Things. He has authored 120 papers in the area of wireless communications.

Xiaodong Lin [F’17] received his Ph.D. degree in information engineering from Beijing University of Posts and Telecommunications, China, and his Ph.D. degree in electrical and computer engineering (with an Outstanding Achievement in Graduate Studies Award) from the University of Waterloo, Canada. He is currently an Associate Professor of Computer Science in the Department of Physics and Computer Science, Wilfrid Laurier University, Canada. His research interests include wireless network security, applied cryptography, computer forensics, software security, and wireless networking and mobile computing. He is a Fellow of the IEEE.