CARAVAN: Providing Location Privacy for VANET

Krishna Sampigethaya*, Leping Huang†, Mingyan Li*, Radha Poovendran*, Kanta Matsuura†, Kaoru Sezaki†
*Department of Electrical Engineering, University of Washington, Seattle, WA 98195-2500
†University of Tokyo, Tokyo, Japan

***Abstract*— In vehicular ad hoc networks (VANET), it is possible to locate and track a vehicle based on its transmissions, during communication with other vehicles or the road-side infrastructure. This type of tracking leads to threats on the location privacy of the vehicle's user. In this paper, we study the problem of providing location privacy in VANET by allowing vehicles to prevent tracking of their broadcast communications. We first, identify the unique characteristics of VANET that must be considered when designing suitable location privacy solutions. Based on these observations, we propose a location privacy scheme called CARAVAN, and evaluate the privacy enhancement achieved under some existing standard constraints of VANET applications, and in the presence of a global adversary.**

I. Introduction

Vehicular ad hoc networks (VANET) enable vehicles to communicate among themselves (V2V communications) and with road-side infrastructure (V2I communications). Such networks present various functionalities in terms of vehicular safety, traffic congestion reduction, and location based service (LBS) applications. Recognizing the potential of VANET, there has been concerted efforts [1], [2], [3] to network vehicles. However, many challenges including the security and privacy issues remain to be addressed [4], [5], [6].

The unique requirements of maintaining liability of vehicles involved in accidents, and ensuring the safety rendered by the communication between vehicles, challenge the network connectivity, privacy, and certain security aspects (discussed later in Section III-D) in VANET. Moreover, advances in localization and tracking techniques enable accurate location estimation and tracking of vehicles in VANET. By tracking a vehicle, it becomes possible to identify the locations visited by the vehicle, thereby, breaching the privacy of the user of the vehicle. Furthermore, the location tracking information about a user can be misused by an adversary. Additionally, identifying the LBS applications accessed by a vehicle, provides private information of the vehicle's user.

In this paper, we address the *problem of allowing any vehicle to be able to achieve unlinkability between two or more of its locations in the presence of tracking by an adversary*. For developing a suitable solution, unlike previous approaches for location privacy in mobile networks (see Section V-C), we account for the constraints posed by vehicular mobility and vehicular applications in VANET (see Section II-D).

Contributions of this paper are the following. (1) We identify that the *group navigation* of vehicles can be used for providing location privacy in VANET. (2) We propose a location privacy scheme called *CARAVAN*, that combines the group navigation with a random silent period enhancement technique, to mitigate tracking of a vehicle. (3) We leverage the group to provide anonymous access to LBS applications, and show when such a solution can preserve a vehicle user's privacy.

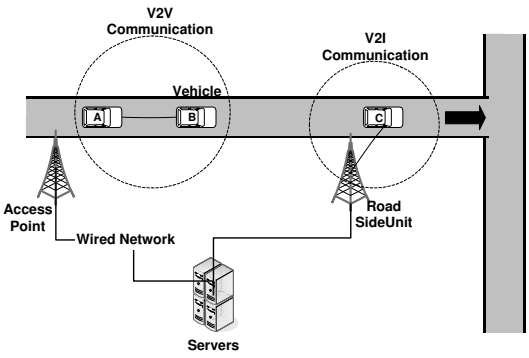

Fig. 1. Illustration of inter-vehicle communication and the components involved. The circles indicate communication between the enclosed nodes.

The rest of the paper is organized as follows. Section II describes the VANET system model and the adversary model considered, and presents the unique constraints of VANET. Section III describes the proposed location privacy enhancement scheme. Section IV evaluates the performance of the proposed solution. Section V covers the related work, and Section VI presents our conclusions.

II. System Model

A. VANET System Model and Assumptions

Fig. 1 illustrates a typical VANET that consists of vehicles, access points on road side, and a collection of location servers. Vehicles move on roads, sharing collective environmental information between themselves, and with the servers via access points.

Fig. 2 illustrates a detailed view of our system model. A vehicle is enabled with on-board communication unit for V2V and V2I communications, and sensor (for example, GPS) and database units to collect environmental information (for example, location, vehicle speed, tire pressure). The communication unit of the access points are called *Road Side Units (RSU)*, which are connected to *location server* by a wired network. The location server records all the *location data* forwarded by the RSUs, and processes the data together with information from other data sources for example, vehicle manufacturers, police, traffic management center, weather information center. The location server also provides an interface for the *location based Service Providers (SP)*. In addition, a trusted *Registration Authority (RA)* provides authentication and authorization service to both vehicles and LBS providers.

As in [2], [5], we assume that a suitable public key infrastructure is available in the VANET. Before joining the

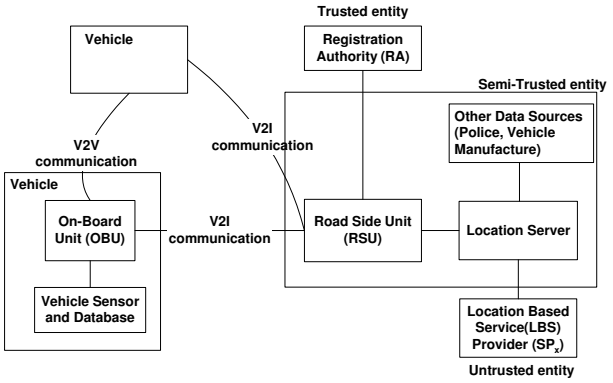

Fig. 2. Illustration of an inter-vehicle communication system.

VANET, each vehicle registers with the trusted RA. We also assume that each LBS service provider registers with the RA, and obtains a public/private key pair. During registration, each vehicle i is pre-loaded with a set of w *pseudonyms* $\{PID_{i,k}\}_{k=1}^{w}$,[1] a public/private key pair $(K_{PID_{i,k}}, K_{PID_{i,k}}^{-1})$, and a corresponding public key certificate $sign_{RA}(K_{PID_{i,k}})$ for each pseudonym $PID_{i,k}$. Each vehicle also registers for any location based service application that is of interest. We assume that only the trusted RA knows the link between the real identity of the vehicle and its associated set of pseudonyms. All communications from a vehicle must contain one of its w pseudonyms as the source address.

B. Trust Assumptions and Adversary Model

We assume that the registration authority (RA) is a trusted entity in our model, as shown in Fig. 2. The infrastructure including the RSUs and the location server are only semi-trusted[2] to operate as expected. We additionally, assume that the RSUs are able to estimate location of a vehicle based on the vehicle's transmission signal.

In our model, we assume a *global passive adversary*. Such an adversary is able to overhear *all* the broadcasts of *all* the vehicles, and hence, able to estimate their locations.

C. Application Scenarios Considered

We consider three typical classes of VANET applications, *cooperative driving*, *probe vehicle data*, and *location based service (LBS)* in this paper. In the *cooperative driving* application, adequate equipped vehicles maintain a very short separation (intra-convoy spacing) between each other and move smoothly with the same pre-defined speed (convoy speed). These vehicles communicate with each other frequently either directly or via communication equipments on road side. For example, in a prototype for cooperative driving in [7], vehicles broadcast their status information (e.g. speed, location, acceleration) every 500 ms. The advantage of cooperative driving is the increase in both safety and highway capacity resulting from the automation and close coordination of vehicles.

[1]The notation used throughout the paper, is in Table II in the Appendix.

[2]A semi-trusted entity operates as expected, but, can still reveal data it obtains during operation.

The *probe vehicle data* represents a class of V2I communication based applications that monitor traffic and road conditions by collecting information from vehicles that are equipped with short range radio (e.g. DSRC, 802.11p) or existing long-range communication devices (e.g. cellular network). Vehicle probe data may include vehicle identity, route segment identity, link time and location, the operational status of the probe vehicle equipment, and any other data that can be measured and communicated by the vehicles. The RSU sends probe data requests over a capture range [8], and vehicles in the capture range reply to the requests. The period between broadcasts of probe replies from vehicles depends on the requirement of applications. For example, according to [9], a typical broadcast interval of probe data for real time congestion estimation is three minutes when probe car volume is 1 vehicle/min.

LBS applications have been proposed for mobile networks. These applications obtain and make use of the most recent location of a mobile node, in order to provide a requested service [10]. For example, the service may be a query by a vehicle to find the nearest shopping mall to its current location.

In the next section, we identify various constraints of vehicular networks that are applicable to the problem addressed in this paper.

D. Relevant Constraints of VANET

VANET poses constraints such as in *mobility of vehicles*, and in safety application requirements. The mobility of vehicles can be observed to have the following unique characteristics: (1) The movement of vehicles is *spatially restricted*. For example, as illustrated in Fig. 1, the movement of vehicles is restricted to be in lanes, in both streets and freeways. (2) The vehicles are *spatially dependent* on each other in movement. For example, as illustrated in Fig. 1, a succeeding vehicle A (following) must keep a minimum *safety distance* [11] from a preceding vehicle B (being followed).

The *safety applications*, as described in Section II-C, impose constraints in terms of the *maximum period between two safety message broadcasts* in cooperative driving, and *maximum period between two replies in probe data*. Therefore, overall, any location privacy enhancement scheme designed for VANET must take into account these unique constraints.

In addition to the above constraints, VANET presents requirements for vehicle liability and safety. In the event of an accident, all the liable vehicles involved need to be verifiably identified. Therefore, to ensure vehicle liability, the safety messages from any vehicle must contain verifiable identification information. Furthermore, for ensuring vehicle safety, the safety messages must be authentic.

III. Proposed Location Privacy Scheme for VANET

In this section, we present CARAVAN, the proposed location privacy scheme for VANET, and describe the enhancement techniques that constitute CARAVAN.

A. Use of Silent Period to Provide Unlinkability Between Locations

In order to achieve unlinkability between two locations a vehicle can simply update its pseudonym. However, as

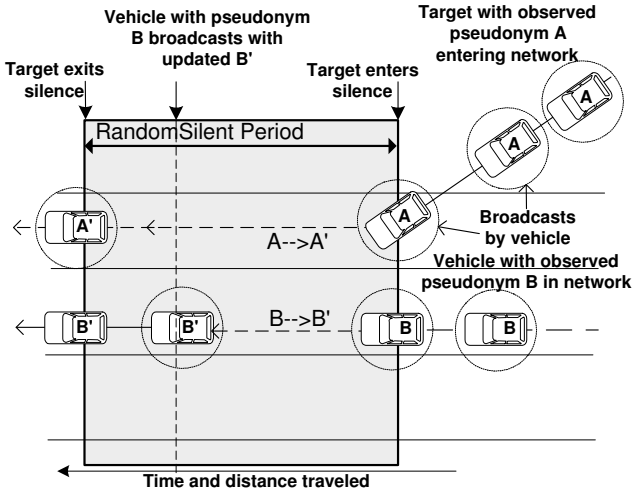

Fig. 3. Illustration of the effect of random silent period when used by a vehicle during network join. A target vehicle entering the network, broadcasts with pseudonym A, and then goes into silence. If a neighboring vehicle updates its pseudonym from B to B' during this silent period, then an adversary can be misled to consider pseudonym B' (and hence, the associated neighbor vehicle's location) to be that of the target vehicle, provided the target vehicle updates to A' before its next broadcast.

observed in [12], despite pseudonym update, it is still possible to link the new and old pseudonyms of a node using temporal and spatial relation between the new and old locations of the node. As a solution the use of a random silent period between update of pseudonyms was proposed in [12]. We make use of *silent period* to provide unlinkability to a vehicle entering the network, by enforcing that the vehicle will remain silent for a randomly chosen period of time.

Fig. 3, illustrates the scenario where a target vehicle enters a network, remains silent, updates its pseudonym from A to A', and broadcasts with A' after a random silent period. If one of the neighboring vehicles also updates is pseudonym from B to B', during this silent period, then the adversary can be misled to track the neighboring vehicle as the target.

However, as discussed in Section II-D, if the vehicles in VANET need to *periodically broadcast* a safety message for cooperative navigation, then the period between safety message broadcasts will be the maximum time between two broadcasts from a vehicle. Therefore, when evaluating the achievable level of anonymity for a vehicle, the time and distance between observations of the vehicle's new and old pseudonyms, must be bounded by this period. Consequently, the maximum silent period will be limited to the fixed value of the period between safety message broadcasts. With only a small and fixed value (order of hundred millisecs) for silent period, it is possible to track vehicles in some cases, based on temporal relation between locations [12]. The achievable anonymity enhancement with constrained values for silent period is evaluated later in Section IV-D.

On the other hand, for VANET applications such as vehicle probe data, that need relatively less frequent broadcasts, we are able to provide a sufficient level of anonymity, by making use of the random silent period technique, as will be shown later in Section IV-E.

B. Use of Group Concept to Avoid Overhearing Pseudonyms

We make the following observations that motivate the group concept applied in our solution.

1) Vehicles in geographical proximity often share redundant information such as road and traffic conditions. Hence, in V2I based applications, such as probe vehicle data, where the vehicles respond to requests received from the infrastructure, not all vehicles need to send replies.
2) As observed in Section II-D, the mobility of vehicles is spatially restricted and spatially dependent. Hence, vehicles in geographical proximity can navigate as a group, with the same average velocity due to the spatial dependency, and with similar direction due to the spatial restrictions, over a period of time.

We make use of the above observations, and propose to enable vehicles to form a *group*. In order to form a group, we restrict the *vehicles to be in a group if each group member can hear broadcasts of every other group member*. Since vehicles in a group will move relative to each other, and on average have the same velocity, a group can be represented by a single vehicle that we refer to as the *group leader*. Then for most of the V2I communication based VANET applications, it is sufficient if only the group leader communicates on behalf of the group. Consequently, the remaining vehicles in the group are able to remain *silent for an extended period* of time that is bounded by the time they remain in the group.

As discussed in the previous section, an extended silent period can enhance the location privacy provided to a vehicle. Therefore, for VANET applications not requiring all vehicles to broadcast, i.e. for applications not requiring very frequent safety message broadcasts from the vehicles, we can increase level of anonymity by employing groups.

We consider the probe vehicle data application, where typically, the vehicles send probe replies once in several tens of seconds. By using vehicular groups, we offer the following benefits: (1) The *silent period* of a group member vehicle is *extended*, if the vehicle does not change group between two probe data requests. (2) Unnecessary *overhead and redundancy* of the neighboring vehicles broadcasting possibly redundant probe data is reduced, since only the group leader replies to the RSU with probe data. (3) A reduced *number of pseudonym updates* (and hence, the number of pseudonyms) are needed to provide the same level of anonymity achieved when the vehicle updates after every broadcast.

However, for safety applications such as cooperative driving, where all vehicles broadcast at a high frequency, the group benefits are not fully realizable. This is because, (1) the extension of silent period is not possible above the safety message broadcast period, (2) each vehicle must broadcast its location, speed, and other spatial parameters for safety, as well as to maintain liability. Hence, under the performance bottleneck of the small safety broadcast period, the advantageous applicability of vehicular group in mitigation of tracking is limited. Nevertheless, vehicular groups can be leveraged to defend against threats on privacy when accessing LBS applications. We describe this advantage of the group below.

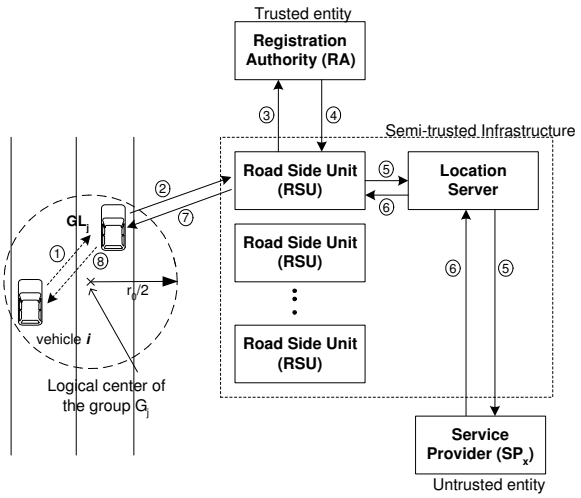

Fig. 4. Illustration of the anonymous access to LBS application provided to a vehicle i which is member of the group G_j with the group leader vehicle being GL_j. The sequence of steps in the protocol are indicated in the figure.

C. Leveraging Group to Provide Unlinkability Between Pseudonym and LBS Application

A global adversary can in certain scenarios, successfully link a vehicle pseudonym with the real identity of the vehicle and hence, its user. For instance, a user might broadcast using a pseudonym, but is located in a geographical area that can be associated with its real identity. When the user accesses an LBS application in such an *identifiable area*, it then becomes possible for the global adversary to identify the LBS application accessed by the user. Such information can lead to the privacy breach of the user.

The use of group, enables us to provide a solution to the above problem by removing the linkability between a vehicle's pseudonym and the LBS application accessed by it. The vehicle accessing the LBS application can make use of the group leader as a *proxy for anonymous access*. We describe this anonymous access protocol below.

1) Protocol description: Fig. 4 shows the anonymous access protocol and the steps involved. Upon receiving the application request from vehicle i (in Step 1), the group leader GL_j of i's group G_j forwards the request with its own address, to the registration authority RA via the RSU (in Step 2-3). The RA validates the application request, and then provides a session key $k_{x,i}$ to both the service provider (SP_x) and vehicle i (Step 4-7). This key is used to encrypt the entire communication that takes place between i and the SP_x. GL_j broadcasts the communication received from SP_x (via RSU) to the group (Step 8).

On termination of the application, the SP_x as well as vehicle i provide the transaction details to the RA, which acts as the arbitrator and resolves any disputes. We note that in order to lower the load of the RA, anonymous payment based protocols such as [13], can be used in the LBS application access. However, we do not provide such a payment scheme here, since it is out of scope of this paper. Due to space constraints, we provide the LBS anonymous access protocol in the Appendix, with the other group protocols.

2) Group Key and Application Address Range: In generating the application request, vehicle i performs the following two steps: (1) randomly chooses an available address A_{aa} from a known *application address range* of the group G_j, (2) broadcasts the application request encrypted with the *group key* k_{G_j} and with A_{aa} as source address. The group key and the address range are obtained by the group members of G_j from GL_j, when joining the group (see Group Join protocol in Appendix). These two parameters *prevent trace back from* GL_j *to* i. Since the *random address* A_{aa} is not associated with vehicle i, the application request from i cannot be associated with any of its pseudonym. This particular feature allows the vehicle i to access the LBS application even in any identifiable area, while also simultaneously broadcasting safety messages with its pseudonym $PID_{i,k}$. The *group key* k_{G_j} on the other hand, prevents tracing i based on the format of application request message that is broadcast to GL_j in Step 1 of the protocol.

Nevertheless, since a global adversary can overhear all broadcasts, it can trace the vehicle i, by relating the location of the overheard application request broadcast sent from i to GL_j, with the more frequent safety message broadcasts by i. Therefore, in order to address this weakness we propose following enhancements by making the group leader GL_j *function as a MIX* [14]. (1) *Removing order of arrival information of the requests.* On receiving application request from i, GL_j waits for one or more requests to be received from other vehicles in the group G_j. The requests are then forwarded to the RSU in a random order (hence, removing the order of arrival information). Therefore, the application app_x accessed by vehicle i cannot be linked to it. However, if all the vehicles access the same app_x then vehicle i can be linked to app_x. (2) *Removing appearance information of the request*. If group key, k_{G_j}, is used to encrypt communications apart from application requests, then the RSU is not able to differentiate the request for app_x based on an encrypted broadcast, from the other group communications. Further, since A_{aa} can be differentiated by the global adversary, to be a new address in the group, only if at least one other group member updates its pseudonym, the tracing of vehicle i can be prevented.

In the following section, we address the different attacks on the proposed scheme, and we suggest suitable solutions.

D. Discussion of Attacks and Solutions for Proposed Scheme

1) Injecting false data: A compromised vehicle in the VANET can misbehave and broadcast incorrect data, with the malicious intent of attacking its neighboring vehicles. However, since each vehicle signs the broadcast safety messages, the identity of any misbehaving vehicle can be verifiably determined. Nevertheless, in order to prevent such attacks on vehicle safety, each vehicle must be able to detect incorrect/malicious safety messages. In [15], a scheme is proposed to enable each vehicle to determine, based on its neighborhood observations, the validity of the data received.

2) Local active attacker: If the group leader colludes with the adversary, then the anonymity of the vehicle accessing the LBS application can be breached under the global adversary

model. For instance, in order to link a vehicle i to the LBS application accessed, a compromised group leader can mix the application requests using an adversary-known deterministic permutation (instead of mixing the requests randomly as described in Section III-C.2). The RSU locates vehicle i from its broadcast to GL_j, and the global adversary upon observing the order of the service providers accessed, can identify that vehicle i is requesting the application app_x from SP_x. We suggest *two* defense mechanisms against attacks by a compromised group leader. For the attack described above, we propose the use of *verification of mixing* to ensure that a random permutation is used by the group leader in mixing the LBS requests. Any verified incorrect mixing will allow the group members (including i) to detect that the GL_j is corrupt. A second defense mechanism is the *group leader rotation protocol* (in Appendix), that restricts attacks by the compromised GL_j to only a certain rotation period. Further, the election of the group leader is randomized to address any collusion between the leader and a group member. Apart from defending against attacks, the leader rotation enables fair provision of privacy to group members, by sharing the leader role amongst them. Due to space limitations, these attacks and defense mechanisms will be analyzed in our future work.

3) Impersonation attack: In the proposed scheme, a vehicle cannot use a random pseudonym, since it must include the associated certificate from the RA in the safety messages (see Cooperative Navigation protocol in Appendix). But, a vehicle may still try to impersonate another vehicle i by using its overheard pseudonym. However, since each vehicle signs the broadcast safety messages, in order to impersonate i the corresponding private key of i must be obtained. Therefore, impersonation attacks can be avoided in VANET. Such defense mechanisms have been considered in [4], [5].

IV. Evaluation of VANET Location Privacy

In this section, we first describe potential tracking methods that can be employed to link two locations of a vehicle.

A. Tracking of Vehicles

1) Simple tracking: In this method, the adversary obtains the target vehicle's location l_{known} and $speed$ at time t, and then estimates, based on possible movement directions, a reachable area A_r around l_{known}, in which the vehicle's actual location l_1 at a future time t_1 can lie. Fig. 5(a), illustrates the simple tracking of a vehicle, and shows the reachable area of the vehicle determined by the achievable speed and silent period ranges.

2) Correlation tracking: As illustrated in Fig. 5(b), in correlation tracking, the adversary uses a vehicles last known location l_{known}, $speed$, and $direction$ at time t to estimate the entity's location l_{est1} at a future time t_1. The estimation is repeated till the maximum silent period is reached.

Note that in both the tracking methods, we assume that the restricted mobility of vehicles prevents them from taking certain directions. Before evaluating the anonymity under the tracking methods by simulation, we first analytically evaluate the level of anonymity that can be achieved under the simple tracking method.

B. Analytical Evaluation of Anonymity

We use two performance measures to evaluate the level of anonymity (unlinkability) achieved in a VANET: (i) the size of the *anonymity set* (ii) the *maximum tracking/identifiable time*. Anonymity set was introduced by Chaum [16], and the size of anonymity set was shown to be a good indicator of how much anonymity is provided. The *anonymity set* of a target, denoted by S_A, is defined as the set of pseudonyms that are indistinguishable from the target pseudonyms to an adversary, and the set includes the target pseudonyms themselves. The size of anonymity set, denoted by $|S_A|$, depends on the knowledge and the tracking method of an adversary. The second measure, *maximum tracking time* of a target, denoted by T_{track}, is defined as the maximum cumulative time that the size of anonymity set of the target remains as one.

We assume that vehicles are uniformly distributed on city streets or freeways with density ρ. Although uniform density neglects the constraints imposed by the street layout, Seskar *et al* [17] showed that uniform distribution is sufficient for estimation of vehicles crossing cell boundaries in mobile cellular networks, when the street layout is not symmetric and the velocities and densities are properly related. In our simulation, we assume that the arrival rate and the departure rate of vehicles are the same. Therefore, the total number of vehicles in the vehicular network deployment region, denoted by N, remains the same statistically, as does the density of vehicles.

Given vehicles are uniformly distributed, the number of vehicles in area A, denoted by $\nu(A)$, distributes according to spatial Poisson process as [18]: $Pr\{\nu(A) = i\} = \frac{(\rho A)^i}{i!} e^{-\rho A}$, with average as ρA.

Suppose that a global adversary is tracking a target by overhearing the broadcast of the target, and is using the *simple tracking method*. The duration between each broadcast can be regarded as silent period, denoted by $speriod$. We first consider the scenario that every vehicle will use a new pseudonym in each broadcast. The reachable area of the target from its last transmission, denoted by A_r, is the half ring bounded by the road/lane layout, as shown in Fig. 5(a). Any vehicle that appears in the reachable region with a *new* pseudonym is a possible candidate for the target to the adversary.[3] Given that there is at least one vehicle, the target, in the reachable region A_r, the probability that the target can be uniquely identified at each transmission, denoted by p_{track}, is:

$$\begin{aligned} p_{track} &= Pr\{\nu(A_r) = 1 | \nu(A_r) \geq 1\} \\ &= \frac{Pr\{\nu(A_r) = 1\}}{1 - Pr\{\nu(A_r) = 0\}} = \frac{\rho A_r e^{(-\rho A_r)}}{1 - e^{-\rho A_r}}. \end{aligned} \quad (1)$$

The expected maximum tracking time is:

$$\begin{aligned} E\{T_{track}\} &= \sum_{i=1}^{\infty} i p_{track}^{i-1} (1 - p_{track}) E\{speriod\} \\ &= \frac{E\{speriod\}}{1 - p_{track}}. \end{aligned} \quad (2)$$

[3]We assume that vehicles periodically broadcast around the same time, then the number of vehicles in the reachable area of the target will be the number of new pseudonyms in its anonymity set. We also note that an adversary cannot distinguish vehicles based on the order of broadcast due to random access.

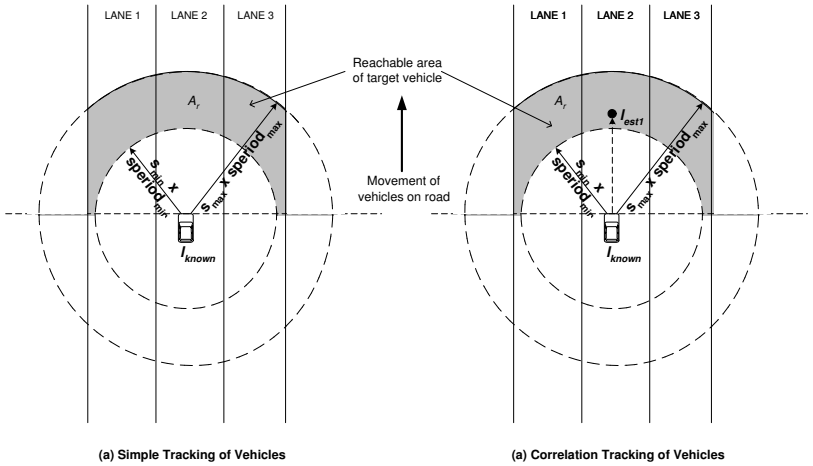

Fig. 5. Illustration of simple tracking and correlation tracking of vehicles. s_{min}, s_{max} are the minimum and maximum speed limits, and, $speriod_{min}, speriod_{max}$ are the minimum and maximum silent period values, respectively. The reachable area is defined by the minimum reachable distance d_{min} and maximum reachable distance d_{max}, where $d_{min} = s_{min} \times speriod_{min}$, $d_{max} = s_{max} \times speriod_{max}$. Location l_{est1} is estimated at time t_1, using the observed velocity of vehicle at last known position l_{known} at time t, where $t_1 \in [t + speriod_{min}, t + speriod_{max}]$. Since vehicles tend to not change direction frequently, they become more susceptible to correlation tracking, as shown in the evaluation.

The expected size of anonymity set of a target is:

$$\begin{aligned} E\{|S_A|\} &= E\{\nu(A_r)|\nu(A_r) \geq 1\} \\ &= \frac{E\{\nu(A_r)\}}{1 - Pr\{\nu(A_r) = 0\}} = \frac{\rho A_r}{1 - e^{-\rho A_r}}. \end{aligned} \quad (3)$$

Next, we consider the case that a vehicle will update its pseudonym with probability $p_u \leq 1$ at each broadcast. In this scenario, the anonymity set of the target equals to l for $l \geq 2$, if and only if (i) the target updates its pseudonym, and (ii) there are $l-1$ other vehicles updating their ID's, out of $\nu(A_r) - 1$ vehicles, which is the number of vehicles in A_r excluding the target. Given the number of vehicles in A_r, the number of vehicles broadcasting with new ID's is Binomial distributed. For $l \geq 2$:

$$\begin{aligned} &Pr\{|S_A| = l\} \\ &= \sum_{i=l}^{N} Pr\{|S_A| = l|\nu(A_r) = i\} Pr\{\nu(A_r) = i|\nu(A_r) \geq 1\} \\ &= \sum_{i=l}^{N} \binom{i-1}{l-1} (p_u)^l (1-p_u)^{(i-l)} \frac{(\rho A_r)^i e^{-\rho A_r}}{i!(1 - e^{-\rho A_r})}. \end{aligned}$$

The probability p_{track}, when the pseudonym update probability of each vehicle is p_u, is:

$$\begin{aligned} &p_{track}(p_u) \\ &= 1 - \sum_{l=2}^{N} Pr\{|S_A| = l\} \quad (4) \\ &= 1 - \sum_{l=2}^{N} \sum_{i=l}^{N} \binom{i-1}{l-1} (p_u)^l (1-p_u)^{(i-l)} \frac{(\rho A_r)^i e^{-\rho A_r}}{i!(1 - e^{-\rho A_r})}. \end{aligned}$$

Then we can apply the above $p_{track}(p_u)$ into Eq. (2) to obtain the expected maximum tracking time.

The average size of an anonymity set is:

$$\begin{aligned} &E\{|S_A| \text{ for given } p_u\} \\ &= \sum_{l=2}^{N} l \cdot Pr\{|S_A| = l\} + 1 \cdot (1 - \sum_{l=2}^{N} Pr\{|S_A| = l\}) \\ &= 1 + \sum_{l=2}^{N} (l-1) Pr\{|S_A| = l\}. \end{aligned} \quad (5)$$

Letting $p_u = 1$, it is easy to verify that Eq. (4) and (5) reduce to Eq. (1) and (3), respectively.

C. Simulation Setup

In order to simulate the mobility of vehicles in vehicular networks, we consider two maps for the vehicles to move: (1) *Freeway*, and (2) *Street* with intersections. For the freeway, we simulate a 4-lane road, with each lane of length 5 km, and with vehicles moving in only one direction. For the street map, we randomly generate a network of intersecting streets on a uniform 2 km $\times$ 2 km grid, with streets separated by 0.5 km. We only consider two types of streets: (a) two lane, one-way, and (b) two-lane, two-way. The lane separation in both the freeway and the street model is 3 meters.

The mobility of vehicles is governed by the following features: (1) *Car following model* [11] which controls the speed and distance of a succeeding vehicle, by making it to keep a *safety distance* (20 meters for freeway, and 10 meters for street) from the preceding vehicle for a certain tolerance time, and then change lane if possible. (2) *Changing lane model*, which allows the vehicle to move to an adjacent lane if there is space in that lane, i.e. if there is no vehicle within safety distance of the position taken when changing lane.

For the street model, we do not account for any intersection behavior, in terms of traffic lights or stop signs. However, at every intersection, we incorporate random mobility by making each vehicle choose to make a left or right turn (if not a one-way street) with probability 0.25 each, or to not change direction with probability 0.5. In both freeway and street

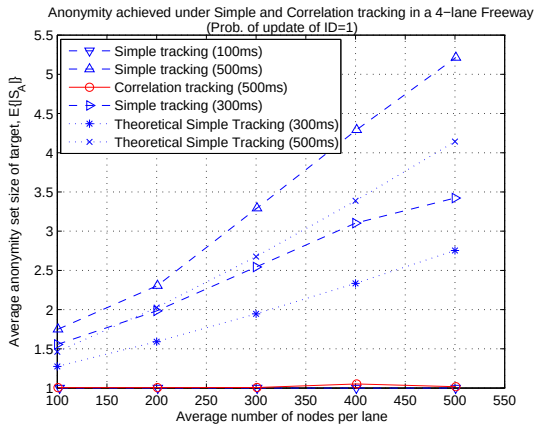

Fig. 6. Average anonymity provided to a target when it updates pseudonym in a 4-lane Freeway, for different number of vehicles (nodes) per lane.

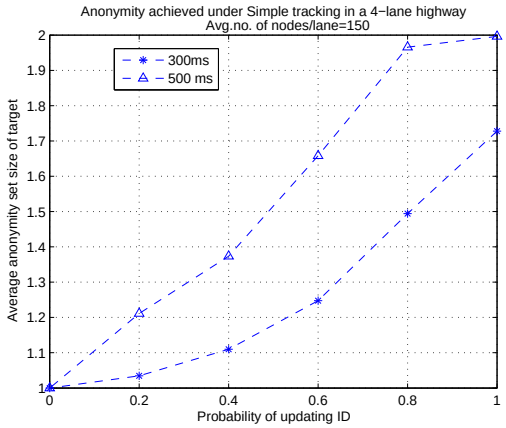

Fig. 7. Average anonymity provided to a target in a 4-lane Freeway, for different probability of updating pseudonym.

models, we do not incorporate the length of vehicles. For the freeway, the speed range is set to $[72$ km/hour, 144 km/hour$]$, and acceleration range is set to $[0$ m/sec^2, 5 m/sec$^2]$. For the streets, the speed range is set to $[36$ km/hour, 72 km/hour$]$, and acceleration range is set to $[0$, 2 m/sec$^2]$.

The traffic volume for freeway is set to 3000 vehicles/hour/lane, and to 1000 vehicles/hour/street for the streets. These numbers are approximated from [19], where 24-hour traffic volume estimates are provided based on real traffic data. At the beginning of the simulation, the vehicles are uniformly distributed in the lanes. It should be noted that due to the higher traffic volume, the average number of vehicles per lane for the freeway is higher compared to the street model. This setting holds under the assumption that there is free flow movement of vehicles, i.e. we do not account for congestion that may arise in streets. Analysis of the anonymity provided for vehicles in real street maps and traffic data will be considered in our future work.

During simulation, for each lane (in freeway map) and each street (in street map), we model the arrival (at pre-determined entry points) and departure of vehicles (at pre-determined exit points) according to Poisson process, based on the traffic volume. The arrival and departure rate are set to be the same, leading to almost same average number of vehicles per lane (street) over time. The border effect of the bounded simulation region on the vehicle mobility, is accounted for by making the vehicle reappear in the region. Currently, we do not integrate any communication traffic model in our simulation.

D. Evaluation of Location Privacy under the Global Passive Adversary Model

We first evaluate the average anonymity a vehicle that can be provided under the global adversary model, where all broadcasts of all the vehicles are overheard by the adversary. Fig. 6, and Fig. 7 shows the average level of anonymity that can be provided when a target vehicle in the freeway, updates its pseudonym between two of its safety message broadcasts. The probability that any vehicle updates its pseudonym, determines how many neighboring vehicles of target change their pseudonym along with the target. Hence, with the decrease

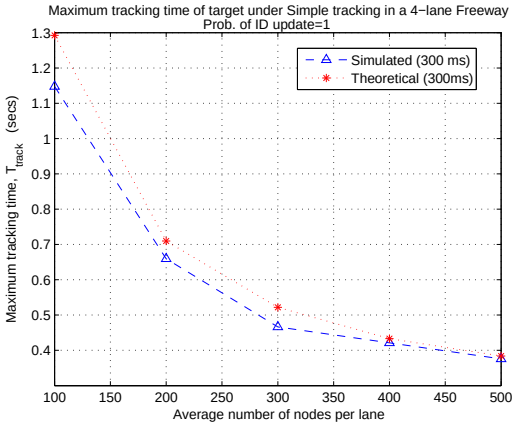

Fig. 8. Maximum Tracking Time of a target in a 4-lane Freeway, for different number of vehicles per lane.

in this probability, it is expected that as in Fig. 7, the target anonymity set reduces to 1. Fig. 8 shows that the maximum tracking time of a target under simple tracking, reduces to the safety broadcast period with increase in number of vehicles per lane. From Fig. 6, 8, we see that the theoretical values for average level of anonymity, and the maximum tracking time, derived from Eq. (3), (2), are slightly pessimistic compared to the simulated values. Fig. 9 shows the achievable anonymity level in the street map. By comparing with Fig. 6, we see that the anonymity level provided in streets is lower. This is due to the relatively lower vehicle density in streets as discussed in the previous section, since we assume a lower traffic volume for streets than for freeways. Due to space limitations, in this paper, we only provide the anonymity enhancement evaluation for freeway model.

It can be observed from Fig. 6, 7, 9 that as we increase the safety message broadcast period from 100 ms to 500ms, the level of anonymity increases under simple tracking. However, *we cannot achieve an increase in the anonymity level under correlation tracking*. Since vehicles tend to not change direction in short time intervals, the correlation tracking method can be used successfully to track them. In order to address this weakness, next, we evaluate the gain in anonymity achieved by increasing the random silent period value.

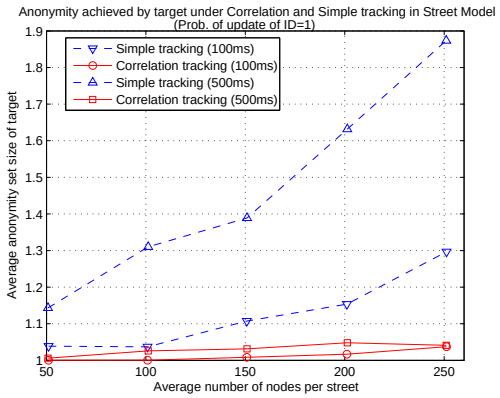

Fig. 9. Average anonymity provided to a target when it updates pseudonym in street model, for different number of vehicles per street.

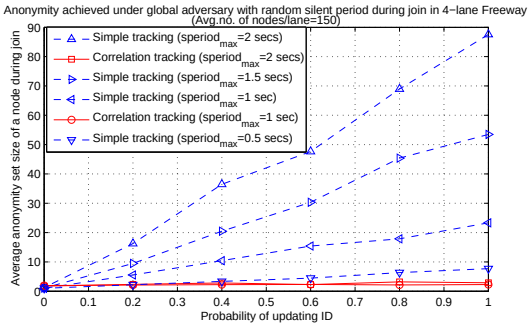

Fig. 10. Enhancement in anonymity obtained from tradeoff of the safety message broadcast period with random silent period during network join.

E. Evaluation of Location Privacy Enhancement with Silent Period

Fig. 10 shows the average anonymity level that can be achieved when a vehicle joining the network remains silent for a random period (less than a maximum value). As we increase the silent period from 500ms to 2 secs, there is a significant increase in the anonymity level under the global adversary using simple tracking. However, we do not achieve a similar gain in the case of correlation tracking. Fig. 11 shows that the silent period has to be increased further to achieve a suitable anonymity level for correlation tracking.

For anonymity under correlation tracking, the vehicles joining the network must remain silent for a period greater than the safety message broadcast period. For instance, from Fig. 11, a vehicle must remain silent and not broadcast any message for at least 1 sec to achieve average anonymity of 2. Hence, for vehicles participating in safety applications, this solution presents a tradeoff between vehicle anonymity and vehicle safety, since by increasing silent period of target beyond the safety message period, we decrease the safety of the target's neighboring vehicles. Therefore, in the following section, we propose another solution for vehicles participating in safety applications. This alternate solution takes into account the observation that the safety message broadcast range for vehicles can be smaller than the broadcast range needed for other VANET applications.

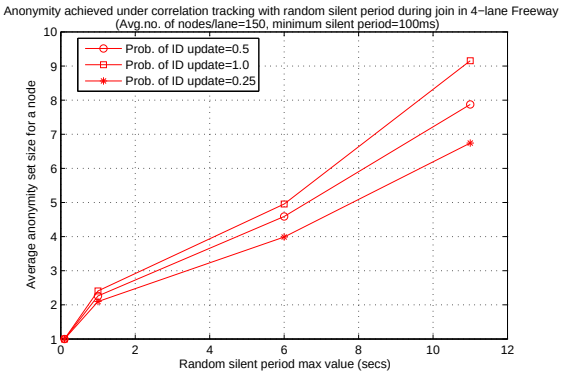

Fig. 11. Enhancement in anonymity obtained under correlation tracking with different values for random silent period during network join.

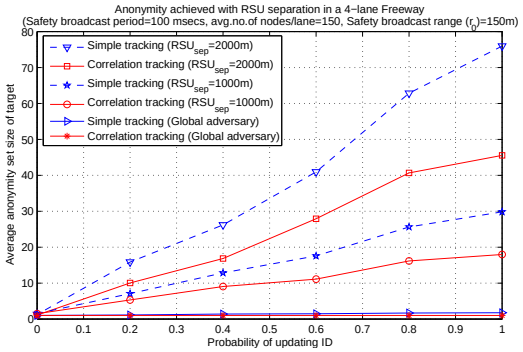

Fig. 12. Enhancement in anonymity obtained from RSU separation.

F. Location Privacy Enhancement with RSU Separation

In [5], an observation is made about the restricted coverage of RSUs due to the separation between them. We illustrate this observation using Fig. 13. Based on the RSU separation (RSU_{sep}), and the safety message broadcast range (r_0), we can define geographical regions called *overheard* and *non-overheard* regions. As seen, in the overheard region, all the safety message broadcasts are received by the RSU. However, the RSUs will not be able to overhear safety message broadcasts of the vehicles in the non-overheard region. We note here that the *vehicles can be assumed to be capable of controlling their transmission range*, and therefore, communicate with the RSU if needed in the non-overheard region. As shown in Fig. 13, the group leader vehicle can increase its transmission power to reply to the probe data request from the RSU.

Given the above scenario, if the target vehicle updates its pseudonym in the non-overheard region, and if there is at least one other vehicle in the non-overheard region that also updates pseudonym, then the adversary may not be able to track the target when it exits the non-overheard region. The anonymity set of the target will include all the vehicles that update their pseudonym along with the target in the non-overheard region.

Fig. 12 shows that with increase in RSU separation, the average anonymity level provided to a target increases significantly under simple tracking, as well as under correlation tracking.

It is worth noting here that by taking the RSU separation into account, we no longer consider tracking under the global

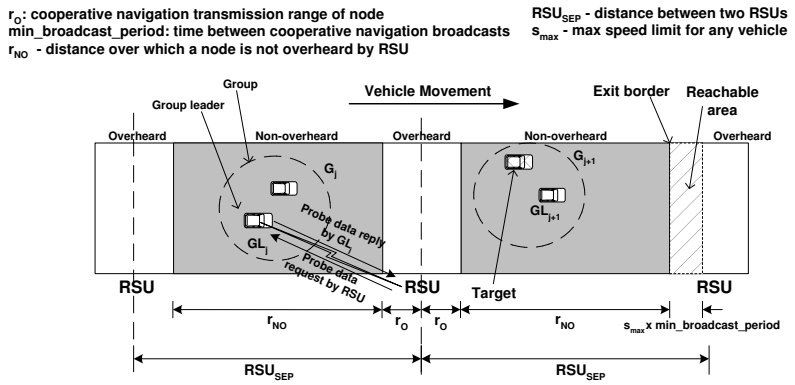

Fig. 13. Illustration of overheard and non-overheard regions in the path of vehicles.

adversary model. The adversary model becomes relatively weaker, since not all the broadcasts are overheard.

G. Comparison of Silent Period with RSU Separation

Comparing the silent period with the RSU separation solution, we see that the two are similar in approach, since both ensure a time period in which the target will move without being overheard. However, comparing Fig. 10 and Fig. 12, it can be seen that the random silent period is unable to provide as much anonymity as RSU separation solution, under correlation tracking. We observe that this is due to the larger time period the target is not overheard in the RSU separation solution. A separation of RSU_{sep} indicates that the time of not being overheard lies in $[RSU_{sep}/s_{max}, RSU_{sep}/s_{min}]$, where s_{min}, s_{max} are the minimum speed and maximum speed, respectively, that the target can assume. Fig. 11 justifies this observation by showing how anonymity is improved with increase in silent period.

On the other hand, the random silent period solution only needs a relatively small time period, to provide equal or better performance under simple tracking, compared to the RSU separation solution. For instance, a silent period of 2 secs achieves the same average anonymity level provided by a RSU separation of 2 km. The reason for this performance difference is that with the RSU separation, due to the known exit border of the non-overheard region, the reachable area of the target is located only at the exit border, and is limited by s_{max} and the minimum broadcast period, as shown in Fig. 13. Hence, even if a vehicle updates more that once in the non-overheard region it will be accounted for only once, i.e. in the reachable area. On the other hand, in random silent period technique, since there are no non-overheard/overheard region assumptions, the reachable area is relatively larger, and hence, if a vehicle updates pseudonym more than once in the reachable area, then it will be accounted for that many times.

V. Related Work

A. VANET Security and Privacy

Security and privacy issues in VANET have just begun to attract attention from both academic and corporate research. Recently, in [3], [4], Hubaux et. al. from EPFL, provide a general framework for security issues in VANET, and analyze in detail, the threats and challenges regarding security and privacy in VANET. They propose several interesting solutions for VANET security such as *Electronic License Plates* (ELPs) that are unique cryptographically verifiable numbers equivalent to traditional license plates, and *location verification* based on verifiable multilateration as an approach to address liability requirements of VANET. Dötzer et. al. [6], [20] from BMW research, have also separately addressed the privacy problems in VANET, and security of V2I communications for safety, particularly between vehicles and traffic lights. In [5], a scheme for providing anonymity in VANET is given, where the vehicles update their keys when changing direction. However, these works do not consider the achievable privacy under global adversary model. In other related VANET security work, Golle et al. [15] address the problem of an adversary injecting malicious data into the network, and propose a general approach to evaluating the validity of the data, where each node searches for possible explanations for the data it has received and collected. ISO/TC204 [21] is responsible for the global standardization activity of ITS. Privacy issue in probe data application is one of the working issues in WG16 of ISO/TC204. However, in comparison with our work, they assume a weaker adversary model. Assuming trusted RSUs not capable of location estimation, they address a policy based approach to protect privacy of users from service providers.

B. Mobility Models for VANET

With emerging interest in VANET, there have been efforts on modeling the mobility of vehicles. In [22], two models (Freeway and Manhattan models) are proposed for mobile ad hoc network simulation. Both of these models account for the spatial dependency between mobile nodes, and restricted movement of nodes in freeway and the street map. Because of their simplicity, we use slight variants of these models in our study, by incorporating additional parameters such as lane changing. The study by Saha and Johnson in [23], accounts for restricted movement on real map data, and uses the current vehicle traffic conditions in determining the path of nodes to their respective destinations. However, they do not take into account the spatial dependency between the nodes. Very recently, the STRAW model has been proposed in [24] that unlike [23], takes into account the spatial dependency between

nodes, but does not incorporate lane changing. In [25], an overview of some existing vehicle traffic simulators is given.

C. Location Privacy Enhancement for Mobile Networks

To protect users from location privacy threats, there are several research studies in mobile networks. Gruteser and his colleagues [10], [19] have worked extensively on protecting location privacy in WLAN. Their approach is based on adjusting the resolution of location along spatial and temporal dimensions, and assumes that nodes provide their location, rather than the location being estimated by any AP/RSU. On the other hand, Beresford [26] proposes the concept of the MIX zone based on Chaum's [14] MIX, to protect location privacy of LBS application users from service providers. The MIX zone for a group of users is a connected geographical region where no application is accessible. Because application providers do not receive any location information when users are in a MIX zone, the user identities are *mixed*. In [12], [27], Huang et. al. propose random silent period to protect user trajectory privacy. However, all of these works assume that the wireless nodes have unrestricted and independent mobility, hence, not considering the unique constraints of VANET.

VI. Conclusions and Future Work

In this paper, we addressed the location privacy threats that arise in VANET due to tracking of vehicles based on their broadcasts, and proposed a solution called CARAVAN. Taking into account the mobility, and the application features in VANET, we identified that by combining neighboring vehicles into groups, it is possible to reduce the number of times a vehicle needed to broadcast for V2I applications such as probe vehicle data. Using group the vehicles can be provided with an extended silent period, which in turn enhances their anonymity. Assuming the global adversary model, and under the safety application constraints of VANET, we evaluated the enhancement of anonymity achieved by our proposed solution. We also suggested an enhancement technique that takes into account the separation between RSUs, and the transmission power control capability of vehicles. Further, we proposed an anonymous access protocol to address threats to privacy that arise due to access to LBS applications, and found that it was robust under the global adversary model, as well as under the safety application constraints. Future work includes evaluation of proposed location privacy solutions under more realistic mobility for vehicles, combined with map data, and with communication traffic models.

Acknowledgment

The authors are grateful to the anonymous reviewers for their valuable feedback on the paper. This work was supported in part by the following grants: NSF grant ANI-0093187-002, and by ARO grant W911NF-05-1-0491. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and should not be interpreted as the views of the National Science Foundation, or the Army Research Office of the U.S. Government.

References

[1] 5.9GHz DSRC. [Online]. Available: http://grouper.ieee.org/groups/scc32/dsrc/index.html

[2] M. E. Zarki, S. Mehrotra, G. Tsudik, and N. Venkatasubramanian, "Security issues in a future vehicular network," in *European Wireless*, 2002.

[3] J.-P. Hubaux, S. Capkun, and J. Luo, "The security and privacy of smart vehicles," *IEEE Security & Privacy*, vol. 2, no. 3, pp. 49–55, 2004.

[4] M. Raya and J.-P. Hubaux, "Security aspects of inter-vehicle communications," in *Swiss Transport Research Conference*, 2005.

[5] ——, "The security of vehicular ad hoc neworks," *Proc. of the ACM Workshop on Security of Ad hoc and Sensor Networks (SASN)*, 2005.

[6] F. Dötzer, "Privacy issues in vehicular ad hoc networks," in *Proc. of the Workshop on Privacy Enhancing Technologies (PET)*, 2005.

[7] R. Hochnadel and M. Gaeta, "A look ahead network (LANET) model for vehicle-to-vehicle communications using DSRC," in *Proc. of the ITS World Congress*, 2003.

[8] ITS probe vehicle techniques. [Online]. Available: http://tti.tamu.edu/documents/FHWA-PL-98-035_c5.pdf

[9] T. Fushiki, T. Yokota, K. Kimita, M. Kumagai, and I. Oda, "Study on density of probe cars sufficient for both level of area coverage and traffic information update cycle," in *Proc. of the ITS World Congress*, 2004.

[10] M. Gruteser and D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," in *Proc. of the ACM MobiSys*, 2003, pp. 31–42.

[11] R. W. Rothery, "Car following models," in *In N.H. Gartner, C. Messer, and A.K. Rathi, editors, Traffic Flow Theory, Chapter 4.*, 2002.

[12] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Enhancing wireless location privacy using silent period," in *IEEE Wireless Communications and Networking Conference (WCNC)*, 2005, pp. 1187–1192.

[13] J. Camenisch and A. Lysyanskaya, "An efficient system for non-transferable anonymous credentials with optional anonymity revocation," in *Advances in Cryptology - EUROCRYPT 2001*, ser. LNCS, vol. 2045. Springer, 2001, pp. 93–118.

[14] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24, no. 2, pp. 84–88, 1981.

[15] P. Golle, D. Greene, and J. Staddon, "Detecting and correcting malicious data in vanets," in *Proc. of the ACM Workshop on Vehicular Ad hoc Networks (VANET)*, 2004, pp. 29–37.

[16] D. Chaum, "The dining cryptographers problem: Unconditional sender and recipient untraceability," *Journal of Cryptology*, vol. 1, pp. 65–75, 1988.

[17] I. Seskar, S. Maric, J. Holtzman, and J. Wasserman, "Rate of location area updates in cellular systems," in *Proc. of the IEEE Vehicular Technology Conference*, 1992, pp. 694–697.

[18] A. M. Mathai, *An Introduction to Geometrical Probability: Distributional Aspects with Applications*. CRC Press, 1999.

[19] M. Gruteser and D. Grunwald, "Enhancing location privacy in wireless LAN through disposable interface identifiers: a quantitative analysis," in *Proc. of the ACM Workshop on Wireless mobile applications and services on WLAN hotspots (WMASH)*, 2003, pp. 46–55.

[20] F. Dötzer, F. Kohlmayer, T. Kosch, and M. Strassberger, "Secure communication for intersection assistance," in *Proc. of the International Workshop on Intelligent Transportation (WIT)*, 2005.

[21] ISO/TC204:transport information and control systems (TICS). [Online]. Available: http://www.sae.org/technicalcommittees/tc204.htm

[22] F. Bai, N. Sadagopan, and A. Helmy, "IMPORTANT: A framework to systematically analyze the impact of mobility on performance of routing protocols for adhoc networks," in *Proc. of the IEEE Infocom*, 2003, pp. 825–835.

[23] A. K. Saha and D. B. Johnson, "Modeling mobility for vehicular ad hoc networks," in *Proc. of the ACM Workshop on Vehicular Ad hoc Networks (VANET)*, 2004, pp. 91–92.

[24] D. R. Choffnes and F. E. Bustamante, "An integrated mobility and traffic model for vehicular wireless networks," in *Proc. of the ACM Workshop on Vehicular Ad hoc Networks (VANET)*, 2005, pp. 69–78.

[25] C. Schroth, F. Dötzer, T. Kosch, B. Ostermaier, and M. Strassberger, "Simulating the traffic effects of vehicle-to-vehicle messaging systems," in *Proc. of the International Conference on ITS Telecommunications*, 2005.

[26] A. R. Beresford, "Location privacy in ubiquitous computing," Ph.D. dissertation, University of Cambridge, November 2004.

[27] L. Huang, K. Matsuura, H. Yamane, and K. Sezaki, "Towards modeling wireless location privacy," in *Proc. of the Workshop on Privacy Enhancing Technologies (PET)*, 2005.

TABLE I
ABBREVIATIONS

GPA	Global Passive Adversary
OBU	On-Board Unit
RSU	Road Side Unit
V2V	Vehicle-to-Vehicle
V2I	Vehicle-to-Infrastructure
IVC	Inter-Vehicle Communication
VANET	Vehicle Ad-hoc NETwork
LBS	Location Based Service
DSRC	Dedicated Short Range Communication

APPENDIX

A. Protocols for Group Formation, Group Join, Group Leave, Group Operation

In the sections below, we detail the various protocols involved in the proposed location privacy scheme for VANETs.

1) Group Join Protocol: Each vehicle (node) i, upon entering the network, periodically broadcasts safety messages for cooperative navigation. However, node i simultaneously attempts to join one of the nearest existing groups. The node i listens for broadcasts from any neighboring group leader GL_j, and then requests GL_j for membership to group G_j. A group leader can be identified by its address included in its broadcasts. The y least significant bits of the group leader's address will be set to zero (see Group Formation protocol). GL_j verifies (using the spatial parameters of i included in the request) if i is in the range of all members of G_j. We restrict the group to have full connectivity, so that group leader rotation is possible. GL_j also verifies the public key of i included in the request, and provides i with the group key k_{G_j} and the LBS application address range, encrypted with public key of i. The pseudocode of the group join protocol is given below.

Group Join Protocol (GROUP_JOIN)

1. i: listen for broadcasts from neighboring group leaders $\mathcal{H}$
if ($|\mathcal{H}| > 0$) and *(waited for* $\leq sp_{max}$*)*
2. i: identify $G_j \in \mathcal{H}$ that was last heard
3. i: change $PID_{i,k-1}$ to $PID_{i,k} \in \{PID_i\}$
4. $i \rightarrow GL_j$:
$request = A_{GL_j}||PID_{i,k-1}||join\_request$
where $join\_request = K_{PID_{i,k-1}}||sign_{RA}(K_{PID_{i,k-1}})$
$||location_i||velocity_i||acceleration_i||timestamp$
5. *if (verified* $K_{PID_{i,k-1}}$*)* and
($location_i$ *is within range of node* a, $\forall a \in G_j$*)*
GL_j: store
$PID_{i,k-1}||K_{PID_{i,k-1}}||sign_{RA}(K_{PID_{i,k-1}})$
$GL_j \rightarrow i$: $reply = PID_{i,k-1}||A_{GL_j}$
$||E_{K_{PID_{i,k-1}}}(k_{G_j}||app\_address\_range)$
else
GL_j: do not reply
endif
6. *if (received reply within* T_{max}*)*
i: set address $A_{i,j} = PID_{i,k}$
i: go to GROUP_OPERATION
else
i: identify $G_k \in \mathcal{H} \backslash G_j$
i: set $G_j = G_k$,
if (less than R_{max} *repetitions without any reply)*
i: go to Step 4
else
i: go to GROUP_FORM
endif
endif
else
i: go to GROUP_FORM
endif

2) Group Formation Protocol: In the above protocol, the node i may not be successful in finding a group to join. The node then creates a group by means of the group formation protocol. i communicates with the RA via the RSU to obtain the group leader ID, GID_j, used in the group leader address A_{GL_j}. This interaction is needed to avoid collision of the group leader addresses, since, y least significant bits of the address are set to be zero, i.e. $A_{GL_j} = GID_j||0^y$. Similarly, collisions in the address range provided for LBS application access is avoided. The pseudocode for the protocol is given below.

Group Formation Protocol (GROUP_FORM)

if (no group heard in GROUP_JOIN) or *(no group leader replied in* GROUP_JOIN)
1. i: choose $PID_{i,k} \in \{PID_i\}$
2. $i \rightarrow RSU$: $leader\_notification =$
$A_{broadcast}||PID_{i,k}||K_{PID_{i,k}}||sign_i(K_{PID_{i,k}})$
3. RSU, RA: verify $K_{PID_{i,k}}$, and generate
$E_{K_{PID_{i,k}}}(GID_j||address\_range)$
4. $RSU \rightarrow i$: broadcast $reply =$
$PID_{i,k}||A_{RSU}||E_{K_{PID_{i,k}}}(GID_j||address\_range)$
5. i: *if (received RSU* $reply$ *within duration* T_{max}*)*
i: generate $A_{GL_j} = GID_j||0^y$
i: go to GROUP_OPERATION, listen for join_request
i: *if (no* GROUP_JOIN *request)* and
(waited for duration W_{max}*)*
i: go to GROUP_JOIN
else
if (number of repetitions of broadcast $< R_{max}$*)*
i: repeat Step 2
else
i: go to GROUP_JOIN
endif
endif
endif

The $address\_range$ in Step 3 is used to provide the random address A_{aa} for the anonymous access to LBS applications. We note that the $address\_range$ can directly generate A_{aa}, or alternatively, it can be used to obtain random y-bit numbers $xx...x$, that can construct the random address $A_{aa} = GID_j||xx...x$.

3) Group Leaving Protocol: The nodes in a VANET are highly mobile, and often a node may accelerate or change direction with time. Consequently, a node can go out of range of the group, thereby leaving its current group, and joining another group near its new location. On the other hand, a node may simply update its pseudonym/address $A_{i,j}$. In either case, the group leader GL_j of node i's current group, must assume that the node has left the group G_j. Therefore in the group leaving protocol, when GL_j does not receive any safety message broadcast with the pseudonym of node i (recorded when joining the group) for a maximum time D_{max}, GL_j assumes that either the node i has left the group or has updated its pseudonym/address $A_{i,j}$. Since in *cooperative navigation*, the nodes periodically broadcast navigational data with period T_n, the group leader can set the period D_{max} to be a multiple of T_n. Node i will self determine if it is out of range of GL_j, and will try to find new group by executing the group join protocol. The pseudocode for group leave protocol is as follows.

Group Leaving Protocol (GROUP_LEAVE)

1. i: compute current distance from group leader GL_j
2. i: *if (going to be out of range from GL_j at $leave\_time$)*
   i: go to GROUP_JOIN
   endif
3. GL_j: *if (no broadcast is received from i for duration D_{max})*
   GL_j: delete entry of $A_{i,j}$ from current group member list
   endif

4) Group Operation Protocol: All the members of the group G_j participate in the group operation protocol, which consists of several subprotocols. The *cooperative navigation protocol* is used for safety applications. In addition, for probe data application, we include an optional *probe data aggregation protocol*, where the group leader aggregates the data received from the members. The aggregated data is included in the reply from the group leader to the RSU probe request in the *probe data collection protocol*. As discussed in Section III-D.2, the group leader node cannot be provided location privacy, since it can be tracked based on its fixed pseudonym/address A_{GL_j}. Hence, periodically the role of the group leader is shared by the group members. This is implemented by the *leader rotation protocol*. The pseudocode for the **group operation protocol** is given below, followed by the various subprotocols.

Group Operation Protocol (GROUP_OPERATION)

1. G_j: go to COOPERATIVE_NAVIGATION
2. *for all* $i \in G_j \backslash GL_j$
   i: listen to broadcast sent by GL_j and go to GROUP_LEAVE
   endfor
3. G_j: optionally go to PROBE_DATA_AGGREGATION
4. GL_j: go to PROBE_DATA_COLLECTION
5. *if (leader rotation is needed)*
   G_j: go to LEADER_ROTATION
   else
   GL_j: go to Step 3.
   endif

In the **probe data aggregation protocol**, only a fraction of p nodes from G_j can broadcast data in each period T_d. The pseudocode for the probe data aggregation between the member of group G_j is as follows. The function $aggregate\_data$ is a suitable spatial data aggregation algorithm, and is not detailed here since it is out of the scope of this paper.

Probe Data Aggregation (PROBE_DATA_AGGREGATION)

1. *for all* $i \in G_j \backslash GL_j$
   $i \rightarrow GL_j$: $PDATA_i = A_{GL_j} || A_{i,j} || location_i || probe\_data_i$ with probability p
   GL_j: record $PDATA_i$
   endfor
2. GL_j: execute $aggregate\_data$ to perform aggregation of all the received $\{PDATA_a\}$ and $PDATA_{GL_j}$, and finally obtain $AGGREGATED\_DATA$
3. G_j: go to Step 1 every T_d

The pseudocode for the **probe data collection protocol** is given below.

Probe Data Collection (PROBE_DATA_COLLECTION)

1. $RSU \rightarrow GL_j$: $probe\_data\_request = A_{broadcast} || A_{RSU} || request\_message$
2. GL_j: *if (no $AGGREGATED\_DATA$)*
   $data = location_{GL_j} || probe\_data_{GL_j}$
   else
   $data = location_{GL_j} || AGGREGATED\_DATA$
   endif
3. $GL_j \rightarrow RSU$: $reply = A_{RSU} || A_{GL_j} || data$

In the Step 2, the group leader checks if there is any data that was aggregated recently. If not, then it broadcasts self generated probe data. We do not specifically detail the $probe\_data$ format in this paper. Note that the $probe\_data\_request$ can include specific data resolution request, i.e. for high resolution aggregated data or for lower resolution group leader only data.

In the **cooperative navigation protocol**, each node independently and periodically broadcasts a safety message every T_n. In order to ensure liability of the message originator, as well as safety of the message receiver, we require each node to sign each safety message, and also include a timestamp to ensure freshness of the message. To enable verification of signature, the node includes the corresponding public key certificate. On receiving a safety

message, node i verifies if the message is valid, and then performs safety computation.

Cooperative Navigation (COOPERATIVE_NAVIGATION)

1. i: $NDATA_{i,j} = A_{broadcast}||A_{i,j}$
   $||sign_i(navigation\_data_i||timestamp)||sign_{RA}(K_{PID_{i,k}})$
2. *for all received* $NDATA_{a,x}$
   i: validate and store $NDATA_{a,x}$
   endfor
3. i: execute $safety\_computation$ using valid $\{NDATA_{a,x}\}$
4. *if (received* $intersection\_RSU\ broadcast = A_{broadcast}||A_{IRSU}||location_{IRSU}$*)*
   i: *if (less than two replies heard)*
   $i \rightarrow intersection\_RSU$: $A_{IRSU}||A_{i,j}||navigation\_data_i$
   endif
5. i: go to step 1 every T_n.

In the above protocol, the data format can be $navigation\_data_i = (location_i, speed_i, acceleration_i, direction_i, timestamp_i)$. Steps 1-3 are used to communicate navigational data between vehicles. The Step 3 is only illustrative of the use of navigational data for safety computation. There may be other applications for such data that is not included here. The algorithm for vehicle safety computation based on the navigational data of neighboring vehicles is out of the scope of this paper.

Step 4 of the protocol, is essentially used to achieve *intersection vehicle collision avoidance* between two groups. To avoid redundancy, not all nodes in G_j need to communicate. On the other hand, due to critical nature of the vehicle collision problem, we need to ensure protocol reliability and vehicle safety. Hence, at least two or more nodes from G_j must communicate with the RSU at the intersection. If we assume that the vehicle (on-board unit) transmission range is relatively smaller than the RSU range, the two or more nodes that reply in Step 4, will be in proximity to the intersection RSU.

As mentioned earlier, in order to provide location privacy for the group leader, it becomes essential to rotate the group leader role (periodically or on demand) among the group members. The following protocol is used to enable the **rotation of the group leader** role in the group G_j.

Group Leader Rotation (LEADER_ROTATION)

1. GL_j: *if (do not want to be group leader) or (end of rotation period)*
   $GL_j \rightarrow G_j$: $notification = A_{broadcast}||A_{GL_j}$
   $||E_{k_{G_j}}\{rotation\_time||leader\_rotation\_notification\}$
2. *forall* $i \in G_j \backslash GL_j$
   i: wait for random time $sp \leq sp_{max}$
   i: mask y least significant bits of $PID_{i,k+1}$, and set the masked $PID_{i,k+1}$ as $A_{GL_{j_{new}}} = GID_{j_{new}}$
   $i \rightarrow G_j$: $reply = A_{broadcast}||A_{i,j}||E_{k_{G_j}}\{leader\_role\_accept||A_{GL_{j_{new}}}\}$
   endfor
3. *if (*GL_j *receives the reply from two or more nodes in* G_j*)*
   GL_j: choose random node i from the nodes that replied
   $GL_j \rightarrow G_j$: $A_{broadcast}||A_{GL_j}||E_{k_{G_j}}\{leader\_role\_granted||A_{i,j}\}$
   else
   if (no reply is received within T_{max}*)*
   GL_j: go to Step 1
   endif
   endif
4. i: broadcast $leader\_notification = A_{broadcast}||A_{GL_{j_{new}}}||PID_{i,k+1}$
5. RSU: verify $leader\_notification$
6. $RSU \rightarrow i$: broadcast ACK if verified to be correct
7. i: *if (not received RSU* ACK *after waiting for* T_{max}*)*
   i: repeat the broadcast in Step 4
   endif

Step 2-3 are used to implement the random election of the new group leader, in order to prevent any attacks that can utilize the knowledge of a deterministic election (discussed earlier Section III-D.2). We can further incorporate a verification mechanism in Step 3, in order to ensure the election of the new leader by the old leader node is truly random.

B. Protocol for Anonymous Access to LBS Application in VANET

Fig. 4 illustrates the scenario, where node i in the group G_j wants to access a location based application offered by service provider SP_x. The steps of the protocol are also illustrated in Fig. 4.

Anonymous Access Protocol (ANONYMOUS_ACCESS)

1. $i \rightarrow GL_j$: $app\_request\_message = A_{GL_j}||A_{aa}||E_{k_{G_j}}\{APP\_REQ\}$
   where $APP\_REQ = app\_request||E_{K_{RA}}(PID_{i,k}||sign_i(PID_{i,k})||h^n(q_i)||app_x)$
2. $GL_j \rightarrow RSU$: $forward\_message = A_{RSU}||A_{GL_j}||location_{GL_j}||APP\_REQ$
3. $RSU \rightarrow RA$: forward APP_REQ
4. RA: compute $MSG = D_{K_{RA}}(E_{K_{RA}}(PID_{i,k}||sign_i(PID_{i,k})||h^n(q_i)||app_x))$
   if (MSG *is not valid)*
   generate $reply = DENY\_REQ$
   endif
   if ($PID_{i,k}$ *in* MSG *is valid)* and
   ($PID_{i,k}$ *has valid access to* app_x*)* and
   ($sign_i(PID_{i,k})||h^n(q_i)$ *in* MSG *is valid)*
   generate $reply = app_x||E_{K_{SP_x}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp))||E_{K_{PID_{i,k}}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp))$
   else
   generate $reply = DENY\_REQ$
   endif

$RA \rightarrow RSU$: $reply$

5. RSU: *if (*$reply == DENY\_REQ$*)*
   go to Step 15
   else
   $RSU \rightarrow SP_x$: send $app\_initiate = location_{GL_j}||E_{K_{SP_x}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp))$
   endif
6. SP_x: *if (received app_initiate from RSU)* and *(able to provide service)*
   compute $D_{K_{SP_x}}(E_{K_{SP_x}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp)))$
   if ($k_{x,i}$ *is valid)* and *(timestamp is not expired)*
   $SP_x \rightarrow RSU$: send $app\_initiate\_response$
   endif
   endif /* $app\_initiate\_response$ is also used to indicate the availability of the SP_x */
7. RSU: *if (received* $app\_initiate\_response$ *within* T_{max1}*)*
   $RSU \rightarrow GL_j$: send $RSU\_response = A_{GL_j}||A_{RSU}||app_x||E_{K_{PID_{i,k}}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp))$
   else
   go to Step 15
   endif
8. GL_j: *if (received* $RSU\_response$ *within* T_{max2}*)*
   $GL_j \rightarrow i$: $app_x||E_{K_{PID_{i,k}}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp))$
   else
   go to Step 15
   endif
9. *for all* i *in* G_j
   if (requested for app_x *access)*
   i: compute $decrypt = D_{K_{PID_{i,k}}}(E_{K_{PID_{i,k}}}(k_{x,i}||sign_{RA}(k_{x,i}, timestamp)))$
   i: *if (successfully obtained* $decrypt$*)*
   if ($k_{x,i}$ *is valid)* and *(timestamp is not expired)*
   i: go to Step 10
   else
   i: go to Step 15
   endif
   else
   i: ignore the broadcast from GL_j
   endif
   endif
   endfor
10. *while (1)* /* two-way communication session between node and SP */
   if (data to be sent to i*)*
   $SP_x \rightarrow RSU$: $E_{k_{x,i}}\{data\}$
   $RSU \rightarrow GL_j$: $A_{GL_j}||A_{RSU}||E_{k_{x,i}}\{data\}$
   $GL_j \rightarrow i$: $E_{k_{x,i}}\{data\}$
   i: decrypt data as $D_{k_{x,i}}\{E_{k_{x,i}}\{data\}\}$
   endif
   i: *if (no data received for* T_{max3}*)* and *(no data to be sent to* SP_x*)*
   go to Step 11
   else
   if (data to be sent to SP_x*)*
   $i \rightarrow GL_j$: $A_{GL_j}||A_{aa}||E_{k_{G_j}}\{app_x||E_{k_{x,i}}\{data\}\}$
   $GL_j \rightarrow RSU$: $A_{RSU}||A_{GL_j}||location_{GL_j}||app_x||E_{k_{x,i}}\{data\}$
   $RSU \rightarrow SP_x$: $location_{GL_j}||E_{k_{x,i}}\{data\}$
   SP_x: decrypt $D_{k_{x,i}}\{E_{k_{x,i}}\{data\}\}$
   endif
   endif
   endwhile
11. $i \rightarrow GL_j$: $A_{GL_j}||A_{aa}||E_{k_{G_j}}\{APP\_FIN\}$ where $APP\_FIN = app_x\_end||E_{K_{RA}}(PID_{i,k}||app_x||k_{x,i}||sign_i(session\_info||timestamp))$
   $GL_j \rightarrow RSU$: $A_{RSU}||A_{GL_j}||location_{GL_j}||APP\_FIN$
   $RSU \rightarrow RA$: forward APP_FIN
12. $SP_x \rightarrow RSU \rightarrow RA$: $SERVICE\_FIN = E_{K_{RA}}(SP_x||app_x||k_{x,i}||sign_{SP_x}(session\_info||timestamp))$
13. RA: *if (received* APP_FIN*)* and *(received* $SERVICE\_FIN$*)*
   RA: $D_{K_{RA}}(E_{K_{RA}}(PID_{i,k}||app_x||k_{x,i}||sign_i(session\_info||timestamp)))$
   RA: $D_{K_{RA}}(E_{K_{RA}}(SP_x||app_x||k_{x,i}||sign_{SP_x}(session\_info||timestamp)))$
   if (decrypted quantities are valid for session between i *and* SP_x*)* and *(*$session\_info$ *in both signatures match)*
   RA: record the decrypted quantities
   go to Step 15
   else
   go to Step 14
   endif
   else
   if (waited for T_{max4}*)* and *(not received* APP_FIN*)* and *(not received* $SERVICE\_FIN$*)*
   go to Step 15
   else
   go to Step 14
   endif
   endif
14. RA, *location server*, i, SP_x: resolve dispute between i and SP_x
15. i, SP_x, GL_j, RSU: terminate session

TABLE II

STANDARD NOTATION USED IN THIS PAPER

Notation	Description
i	A entity/node in the VANET.
$i \rightarrow j$	Entity i broadcasts to entity j.
G_j	A group j of nodes in the VANET.
$\mathcal{N}$	Set of all n nodes in the VANET, i.e. $\|\mathcal{N}\| = N$.
$\mathcal{G}$	Set of all g groups in the VANET, i.e. $\|\mathcal{G}\| = g$.
$\mathcal{H}$	Set of groups in the VANET. $\mathcal{H} \subseteq \mathcal{G}$.
L_{max}	Maximum size for a group.
GL_j	Group Leader of group G_j.
GID_j	Group ID of group G_j.
$PID_{i,k}$	k^{th} pseudonym of node i. Each node i has a set of w pseudonyms, $\{PID_{i,k}\}_{k=1}^{w} = \{PID_i\}$.
A_{GL_j}	ID of GL_j. Note that $A_{GL_j} = GID_j \|\| 0^y$, where y is size (in bits) of node ID field.
A_{aa_j}	LBS application access address selected from an address range for group G_j.
$A_{i,j}$	ID of node i that is a member of group G_j. Note that $A_{i,j} = PID_{i,k}$ or $A_{i,j} = GID_j \|\| A_{aa_j}$.
$A_{broadcast}$	Broadcast address for network.
$A_d \|\| A_s \|\| data$	Destination address || Source Address || Data.
$speriod$	Random silent period. $speriod_{min} \leq speriod \leq speriod_{max}$.
s_{min}, s_{max}	Minimum and maximum speed limits for a node.
R_{max}	Maximum number of broadcast repetitions.
T_{max}	Maximum waiting period for an ACK or a reply.
W_{max}	Maximum waiting period for a group join request.
$x \|\| y$ or (x, y)	x concatenated to y.
$\{x\}$	A set of elements.
/** comment **/	Comments in the pseudocode.
K_x, K_x^{-1}	Public and private key pair of entity x.
$k_{x,y}$	Pairwise symmetric key of two entities x, y.
k_{G_j}	Symmetric key of group G_j.
$c = E_{K_x}(m)$	Encryption of message m with public key K_x.
$D_{K_x}(c)$	Decryption of ciphertext c with private key K_x^{-1}.
$E_{k_x}\{.\}$, $D_{k_x}\{.\}$	Encryption and Decryption with symmetric key k_x.
$sign_i(m)$	Digital signature on message m with private key of entity i.
$h(m)$	Cryptographic hash of a message m. Also, $h^n(m) = h(h^{n-1}(m))$, $n \geq 2$.
q_i	A secret quantity of node i.