A Privacy-Preserving Scheme for Location Based Services in the Internet of Vehicles

Jiaqi Huang, Yi Qian, Rose Qingyang Hu

***Abstract*—Ubiquitous information exchange is achieved among connected vehicles through the increasingly smart environment. The concept of conventional vehicular ad hoc network is gradually transformed into the Internet of Vehicles. Meanwhile, more and more location-based services (LBSs) are created to provide convenience for drivers. However, the frequently updated location information sent to the LBS server also puts user location privacy at risk. Thus, preserve user location privacy while allowing vehicles to have high-quality LBSs is a critical issue. Many solutions have been proposed in the literature to preserve location privacy. However, most of them cannot provide real-time LBS with accurate location updates. In this paper, we propose a novel location privacy-preserving scheme, which allows vehicles to send accurate real-time location information to the LBS server while preventing being tracked by attackers. In the proposed scheme, a vehicle utilizes the location information of selected shadow vehicles, whose route diverge from the requester, to generate multiple virtual trajectories to the LBS server so as to mislead attackers. Simulation results show that our proposed scheme achieves a high privacy-preserving level and outperforms other state-of-the-art schemes in terms of location entropy and tracking success ratio.**

***Keywords*—location privacy, location based service, internet of vehicles, vehicular networks**

I. Introduction

Vehicular ad hoc networks (VANETs), which allow vehicles on the road to connect with each other to share information, have been studied for many years. Recently, vehicles are increasingly equipped with smart devices, which allow vehicles access to ubiquitous information of the surrounding environment as well as the internet. The original concept of VANETs is gradually upgraded into the Internet of Vehicles (IoV), which consists of inter-vehicular network, intra-vehicular network, and vehicular mobile Internet[1]. As the major component of Intelligent Transportation System, IoV targets to help traffic management, improve road safety, provide infotainment, etc. With the fast development of positioning technologies such as the Global Position System (GPS), various location-based services (LBSs) have emerged in the IoV paradigm. With LBSs, vehicle users can obtain useful spatial information by uploading their physical locations to the LBS server, thus greatly improve convenience. For example, a driver can obtain continuously updated traffic conditions so as to find the best route or find the availability of nearby parking lots, restaurants, and gas stations.

Since the precondition of using LBSs is uploading a user's physical locations to the LBS server and the LBS server is usually an untrusted third party, protecting the location privacy from being leaked to attackers is of great concern[2]. Many pseudonym-based schemes have been proposed to preserve security and privacy in vehicular communications. However, pseudonyms alone cannot guarantee location privacy[3]. Attackers may link anonymous messages that belong to the same sender by analyzing the location, speed, direction information and finding the correlations. Thus, extra mechanisms besides pseudonyms need to be implemented to preserve location privacy.

Various solutions have been proposed to preserve location privacy in IoV. Major ideas of these schemes can be summarized as k-anonymity, silent period, mixed zones, caching, and virtual paths. The k-anonymity scheme employees k locations in the LBS request to form a cloak region for the user so that the attacker can only know a rough area instead of the accurate location of the user. The silent period scheme forbids message transmission in some period of time or within some areas, thus the attacker will lose track. The mixed zones scheme asks a group of vehicles to change their pseudonyms at a specific time and location to prevent attackers from finding correlations of used and new pseudonyms. The caching scheme suggests storing solutions of LBS requests in vehi-

Manuscript received ******, 2021; revised ******, 2021; accepted ******, 2021. This work was supported by the National Science Foundation under grants CNS-2007995 and CNS-2008145. The associate editor coordinating the review of this paper and approving it for publication was *******.

J. Q. Huang. University of Central Missouri, 116 W South St, Warrensburg, MO 64093, USA (e-mail: jhuang@ucmo.edu).

Y. Qian. Unviersity of Nebraska-Lincoln, 1110 S 67th St, Omaha, NE 68182, USA (e-mail: yi.qian@unl.edu).

R. Q. Y. Hu. Utah State Unviersity, Logan, UT 84322, USA (e-mail: rose.hu@usu.edu).

Tab. 1 Comparisons of existing privacy-preserving schemes

Schemes	Main Strategy	Degree of location privacy protection	Number of generated beacons	Real-time service	Accurate location update	Disadvantages
K-anonymity	Send k LBS requests together to the LBS server so as to hide user's real location among k.	Medium	Medium	No	Yes	The performance is poor in low vehicle density scenario. In some Schemes, a trusted third party is required, which has the single point of failure problem.
Silent Period	Vehicles do not send out any messages to keep silent under some situations.	High	Low	No	No	Vehicles cannot have any LBS during the silent period which is undesirable especially for safety-related applications.
Mixed Zones	Vehicles refresh their pseudonyms at some specific mixed zones so as to prevent attacker link the new pseudonym with the old one.	Medium	High	Yes	Yes	The performance is low if the number of vehicles gathered at the mixed zone is small. If a vehicle does not pass any mixed zone, then there is no protection.
Caching	Cache some results of previous LBS requests in vehicles or access points near the road so as to reduce the number of requests that sent to the server.	Low	Low	No	Yes	Cached results cannot always satisfy vehicle user's demand. If a new request needs to be sent, then the location privacy is still at risk. So the caching method should be applied with other protection mechanisms.
Virtual Path	Generate multiple virtual paths during the trip so that the LBS server can only guess the vehicle's real location from a bunch of plausible destinations.	High	Medium	Yes	Yes	If all the virtual paths are carefully analyzed by attackers, some virtual paths can be manually removed, which increases the tracking success ratio.

cles or roadside units. If a vehicle can find the solution from its own database or the nearby roadside unit, then there is no need to send an LBS request to the LBS server, which reduces the risk of location leakage. However, these schemes cannot provide real-time LBS with accurate location information. Thus, Lim et al.[4] proposed a scheme that using virtual routes to overcome this challenge. In Ref. [4], a vehicle may generate multiple fake routes that leading to different destinations during the trip. The vehicle sends both the faked location and its real location to the LBS server and only takes the response of its real location. After some time of traveling, plenty of plausible fake locations will be generated to confuse the attacker. However, Ref. [4] is only designed for the LBS system that does not use pseudonyms, which may not be suitable for many applications. Another shortcoming of Ref. [4] is that it only considers the interaction between two vehicles at a time, which limits the performance. Moreover, an agreement needs to be established between the two vehicles, making extra communication overhead. In Summary, preserving user's location privacy is a challenging task. If users send accurate location information continuously to the LBS server, their location history can be revealed by analyzing the accumulated path information. If users send ambiguous location information or send location information with a lower frequency to preserve the location privacy, the quality of service could be low, or sometimes even no service at all. Moreover, due to the physical restrictions of vehicles, if "fake locations" are not properly generated, they can be easily filtered by attackers and the real location history can be easily recognized.

Considering the limitation of the existing location privacy-preserving schemes, in this paper we propose a novel scheme. The proposed scheme allows multiple vehicles to interact with each other so that a higher privacy level can be achieved. Moreover, the proposed scheme can preserve location privacy while having real-time LBS without any degradation of location accuracy. Simulation results exhibit that the proposed scheme achieves better performance than other schemes in terms of location entropy and tracking success ratio.

The rest of this paper is organized as follows. Section II discusses the related work in location privacy-preserving. Section III illustrates the LBS system model, threat model, and design objectives of this paper. Section IV describes the proposed scheme in detail. Section V evaluates the performance of the proposed scheme and compares it with other related schemes. Section VI concludes the paper.

II. Related Work

Preserving location privacy for LBS is critical to protect user's privacy and safety. Many solutions have been proposed to preserve location privacy in the IoV environment. Most of the existing solutions can be categorized into k-anonymity, silent period, mixed zones, caching, and virtual paths.

In k-anonymity schemes[5-9], usually a group or a clock region that contains k vehicles is formed before vehicles send LBS requests. After that, k LBS requests sent from the vehicles inside the group are collected and delivered to the LBS server together. In this way, the LBS server cannot distinguish the location of a target user from other $k-1$ requests. The privacy level provided by k-anonymity is closely related to the number of vehicles in the formed group or the clock region, i.e. the value of k. An improved version of k-anonymity was proposed in Ref. [10], which is named CliqueCloak. In that scheme, the k-anonymity is personalized since users can adjust the level of privacy by themselves. However, the service availability of k-anonymity schemes is low in the suburban area, where the vehicle density is low. Because the LBS request will be rejected if the number of collected requests is less than k. Moreover, there is an extra delay when waiting for all the requests to be collected.

In silent period schemes[11-13], vehicles do not send any message in a certain period of time or within a certain area. As long as they are out of the silent period, they start to send messages again but with new pseudonyms. In this way, the attackers cannot link the new pseudonym to the previous one by tracking the continuous location update. The disadvantage of this kind of scheme is that they cannot provide real-time service, since no service can be provided during the silent period.

In mixed zones schemes[14-17], vehicles are recommended to change their pseudonyms at public social spots, where a large number of vehicles may get together, e.g. crossroads and parking lots. In this kind of scheme, each vehicle always holds a bunch of valid pseudonyms during the trip. Pseudonyms are not changed when they are overly used but at social spots. For example, a group of vehicles stop at a crossroad due to a red light. Then, all these vehicles change their pseudonyms when the traffic light turns green. In this way, the attacker cannot track a vehicle by analyzing location information since all the location information is the same at that moment. The performance of the mixed zone scheme is closely related to the number of vehicles gathered at the social spot. However, social spots with a large number of gathered vehicles are not common. To overcome this issue, a new scheme[18] is proposed that vehicles can form the mixed zone dynamically during the trip, e.g. when a group of vehicles are moving at a similar speed. However, the performance of these schemes is not satisfactory when the trip time is short.

Caching based schemes[19-21] are very different from the above mentioned schemes. The main idea of caching-based schemes is to reduce the number of requests sent to the LBS server so that the LBS server does not have enough information to conduct attacks. In Ref. [19], each vehicle stores all the solutions of their requests in the local database. If the solution can be found in the local database, then the vehicle will not send the request to the server. Another way to use caching is to store the LBS solutions at roadside units and let roadside units broadcast the solutions periodically. Vehicles do not need to send new requests if they can obtain the solution from roadside units directly, thus reduce the risk of location leakage. However, location privacy is still at risk when the solution of an LBS request has not been cached yet. Moreover, the caching-based schemes cannot provide real-time services.

In virtual path schemes, instead of generating multiple dummy locations to hide the real location, these schemes focus on generating fake paths to mislead attackers. Lim *et al.* proposed a Mutual Obfuscating Paths (MOP) scheme to provide real-time LBS while preserving location privacy[4]. In the MOP scheme, a vehicle A uses the location information transmitted through Dedicated Short-Range Communication (DSRC)[22] beacon to predict future paths of surrounding vehicles. If its own path will converge with another vehicle B within time β, then it sends a request to the vehicle B to make an agreement. If the agreement is established between vehicles A and B, then the vehicle A will generate a virtual route using vehicle B's current location as the origin and the vehicle A's real location after time β as the destination. And the vehicle B performs the same process as the vehicle A does. After a trip time of 10 minutes, multiple fake paths will be generated by vehicle A with a large number of faked destinations. Thus, the attacker can hardly identify the real route of the vehicle from all possible routes. Cui *et al.* proposed a Privacy-Preserving scheme for Real-time Location Data (PPRLD), which has a similar idea to MOP[23]. In PPRLD, there is no requirement of establishing an agreement between two vehicles, thus has lower communication overhead and is more flexible. The authors claim that the PPRLD scheme achieves a higher privacy level than the MOP scheme.

A recent study[24] summarized a statistical comparison of some location privacy-preserving schemes. Comparisons are based on extensive simulation results obtained from PREXT[25], which is an extension of Veins[26]. Ref. [24] points out that the silent period based schemes provide the lowest traceability with relatively low overhead. However, the low traceability provided by the silent period scheme is obtained by not having any service during a great portion of the trip. In this paper, we summarize functional comparisons of the existing location privacy-preserving schemes in Tab. 1. The table also shows the disadvantages of silent period schemes are that they can not provide real-time service

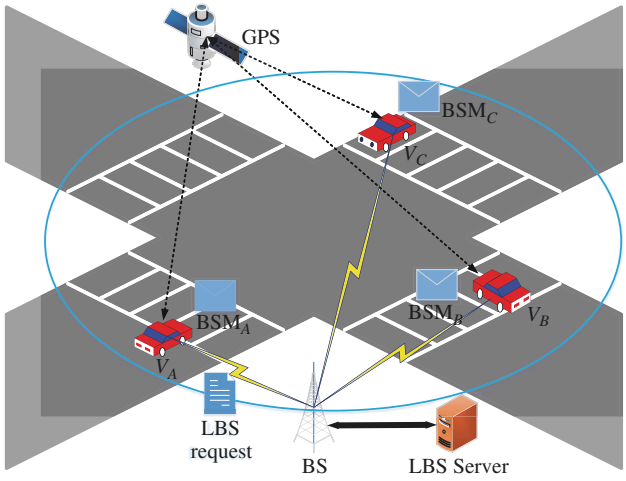

Fig. 1 A general LBS system

nor accurate location updates. From the table, we can see that only the virtual path schemes (MOP and PPRLD) can provide real-time LBS service with accurate location updates. However, the MOP scheme is designed for the LBS systems that do not use pseudonyms. If pseudonyms are used in MOP, the real route and the virtual route can be easily distinguished by following the same pseudonyms. The potential drawback of the PPRLD scheme is that all the generated virtual routes are converged with the real route of a vehicle. Thus, it is highly possible that the longest route is the real route of a vehicle, which might be easily identified by the attacker. Both the MOP scheme and the PPRLD scheme only involve two vehicles at a time. A virtual path scheme that involves multiple vehicles to preserve location privacy has not been studied. In this paper, we propose a novel location privacy-preserving scheme to prevent a user from being tracked by the attacker while having LBS, which also provides real-time service and accurate location updates. Moreover, the proposed scheme can involve multiple vehicles to generate virtual paths so that a higher level of location privacy can be achieved.

III. System Model

In this section, we describe the LBS system, threat model, and design objectives that considered in this paper.

A. The LBS System Model

We consider a general LBS system model that consists of vehicles, base stations (BSs), and a third party LBS server, as shown in Fig. 1. All vehicles traveling on the roads have a GPS receiver and they can receive real-time location information from the GPS. For the security concern, each vehicle broadcasts basic safety messages (BSMs) periodically. The BSM contains the vehicle's location, speed, and direction information, and the broadcasting period is usually smaller than 300 ms. Vehicles are connected to the internet using the cellular networks through BSs. If a vehicle needs to have LBS, a series of LBS requests should be sent periodically to the LBS server. The period of sending LBS requests, which is different from BSMs, is typically set to 1 to 3 seconds. In this paper, we consider an LBS request has the general format of $\{PID, Loc, T, \sigma\}$, where PID, Loc, T, and σ represent the pseudo identity, current location, time stamp, and signature, respectively. The PID is used to protect the user's identity privacy, and the signature is used to provide other security services for the request, e.g. authentication, message integrity, and non-repudiation. After receiving an LBS request, the third-party LBS server will provide service according to the location information provided in the request.

B. Threat Model

We consider that the BSs and the LBS server are untrusted. Moreover, anyone that has access to the LBS server database is also considered untrusted. These attackers try to track LBS users by following the periodically updated LBS requests. Although the PID can protect user's privacy to some extent, the LBS requests are like pins that puncture user's locations point by point. The compromised location privacy may further violate the user's identity privacy. For example, if the attacker finds the destination of a user is a residential address, then the real identity of the user may be identified. Thus, implementing location privacy-preserving mechanisms besides using the pseudo identity is necessary. In this paper, we consider the attacker has full access to the LBS server's database and can track a vehicle by cross-referring the location information contained in the LBS requests. However, we consider all attackers are passive attackers that will not modify the messages transmitted in the network. We also do not concern with the security issues of the PID and the σ, since many pseudonym based security schemes can be found in the literature, e.g., Refs. [27,28].

C. Design Objectives

Our design objectives are stated as follows. First, a user can have real-time LBS with negligible delay. Second, the updated location information is accurate instead of a cloak region or a piece of vague location information, thus the provided LBS has high quality. Third, the location privacy is well preserved even the location information is frequently updated. In other words, the attacker can only find the real location of a user with a negligible probability. Designing a privacy-preserving scheme that satisfies the above-mentioned properties remains a big challenge. Thus, in this paper, we propose a novel location privacy-preserving scheme to address this problem.

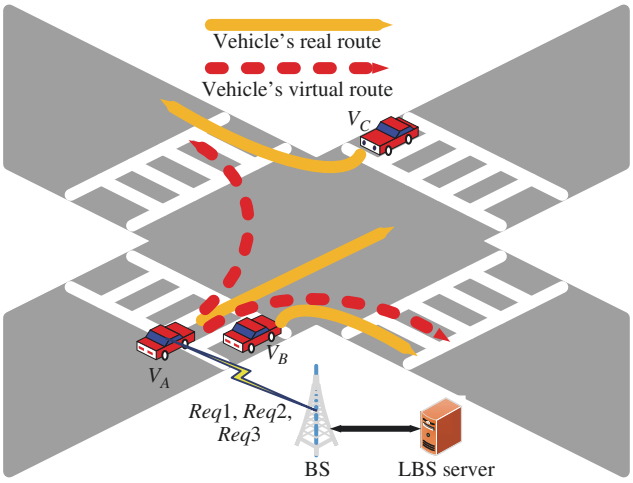

Fig. 2 Overview of the proposed scheme

Tab. 2 Symbol notations

Symbol	Notation
V_i	vehicle i
PID_i	pseudonym of vehicle i
T	current time stamp
δ	time interval of sending LBS requests
σ	signature
λ	level of divergence
T^{-1}	time stamp of the moment that δ second before T, i.e. $T^{-1} = T - \delta$
$Loc(V_i, T)$	location of V_i at time T (in this paper, we consider the location is presented as a vector)
$SP(V_i, T)$	speed of V_i at time T
SP_{max}	maximum speed limit
SD_i	V_i's set of candidate shadow vehicles
$\overrightarrow{Dir}(V_i, T)$	driving direction of V_i at time T
acc_{V_i}	acceleration of V_i
CR	maximum communication range of vehicles through DSRC
$Dist_{max}$	maximum distance a vehicle can travel during the interval of two consequent LBS requests.
$Dist_{thld}$	maximum distance to replace the generated virtual location of a vehicle with the real location of a shadow vehicle without raising suspicion
$Dist(V_i, V_j, T)$	Euclidean Distance between V_i and V_j at time T

IV. The Proposed Scheme

In this section, we describe the proposed location privacy-preserving scheme in detail. Notations of used symbols in our proposed scheme are listed in Tab. 2.

A. Overview

In the proposed scheme, vehicles utilize the information in the BSMs received from nearby vehicles to create virtual routes. As shown in Fig. 2, vehicle A, B, and C (denoted as V_A, V_B, and V_C) are traveling along the road while using LBSs. Meanwhile, Each vehicle keeps receiving the BSMs and be aware of the location, speed, and direction of surrounding vehicles. In this paper, we consider all vehicles that within the communication range of DSRC are surrounding vehicles. Based on traffic information obtained from BSMs, a vehicle always monitors surrounding vehicles and selects some vehicles as its shadows to generate virtual routes. In this example, the V_A selects V_B and V_C as its shadows. Next time when V_A needs to send LBS request, it will send multiple LBS requests to the LBS server instead of a single request, e.g. $Req1 : \{PID_A, Loc(V_A, T), T, \sigma_A^1\}$, $Req2 : \{PID_A, Loc(V_A^B, T), T, \sigma_A^2\}$, and $Req3 : \{PID_A, Loc(V_A^C, T), T, \sigma_A^3\}$. Meanwhile, V_B and V_C will also send multiple LBS requests to the LBS server in a reciprocal way. For example, the V_B will send $Req1' : \{PID_B, Loc(V_B, T), T, \sigma_B^1\}$, $Req2' : \{PID_B, Loc(V_B^A, T), T, \sigma_B^2\}$, and $Req3' : \{PID_B, Loc(V_B^C, T), T, \sigma_B^3\}$. Since all these LBS requests have valid signatures, the LBS server will reply to each vehicle with multiple responses corresponding to all different locations. Each vehicle only uses the response based on its real location, and discards all other responses. V_A will keep using the location of V_B and V_C to send redundant LBS requests until V_A cannot get BSMs from V_B and V_C anymore (V_B and V_C are beyond the communication range of V_A). To prevent the attacker from identifying the real route of V_A by following its PID, once V_B or V_C is out of the communication range of V_A, the V_A changes its PID. Note that the communication range of vehicles is considered the same. So, when V_A is beyond the communication range of V_B, V_A and V_B will change their PID simultaneously. In this way, the attacker cannot determine which route belongs to V_A or V_B. The larger the number of selected shadow vehicles, the lower the probability that the attacker can track a vehicle successfully. A flowchart of the proposed scheme is shown in Fig. 3.

B. Shadow Selection

The selection of a shadow is based on the position and driving direction of a vehicle. The main idea is to select a surrounding vehicle that will converge at some moment and then diverges. The vehicle V_A keeps receiving the BSMs and obtains basic information of surrounding vehicles, e.g. $Loc(V_B, T)$, and $Loc(V_C, T)$. Then V_A calculates the Euclidean distance between itself and each of the surrounding vehicles, e.g. $Dist(V_A, V_B, T)$. The vehicle V_A consider V_B as a candidate for being a shadow if

$$Dist(V_A, V_B, T) < Dist_{max}, \quad (1)$$

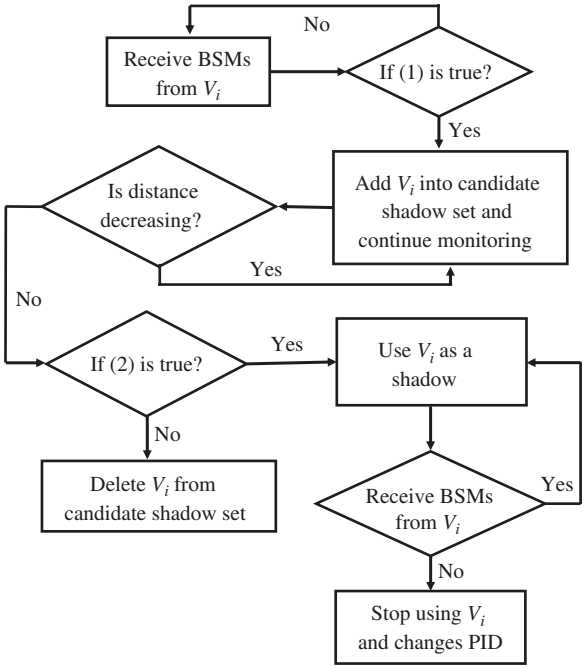

Fig. 3 The flowchart of the proposed scheme

Algorithm 1: Pseudo Code of the Shadow Selection

```
while Sending LBS requests do
    Collecting BSMs from surrounding vehicles;
    Calculate and record the distance between V_A and
     V_i at time T as Dist(V_A, V_i, T);
    if Dist(V_A, V_i, T) >
     Dist(V_A, V_i, T^-1) && Dist(V_A, V_i, T) <
     Dist_max && Dir_A · Dir_B < λ then
        Put V_i into the set of V_A's shadow vehicles
         SD_A, SD_A = SD_A + V_i;
    end
    for All V_i ⊂ SD_A do
        if Dist(V_A, V_i, T) < CR then
            Prepare a LBS request with V_i's location;
        else
            Change PID;
            Delete V_i from V_A's set of shadow vehicles
             SD_A = SD_A − V_i;
        end
    end
    Prepare a LBS request with V_A's current location;
    Send all prepared LBS requests to LBS server;
end
while Receiving LBS responses do
    Take the response with V_A's real location
     information and discard all other responses;
end
```

where $Dist_{max}$ is the maximum distance a vehicle can travel during the interval of two consequent LBS requests. During the trip of V_A, multiple vehicles may converge with V_A at some specific time slots, then they are considered as candidates of V_A's shadows. V_A keeps updating the distance between itself and each vehicle in the candidate set. If the distance is becoming smaller, V_A does nothing but keep updating. If the distance is becoming larger, which means the candidate vehicle is going to diverge from V_A, then V_A checks whether the future path of the candidate vehicle diverges from V_A's real path as follows.

$$\overrightarrow{Dir}(V_A,T)\cdot\overrightarrow{Dir}(V_B,T)<\lambda, \tag{2}$$

where the $\overrightarrow{Dir}(V_A,T)$ and $\overrightarrow{Dir}(V_B,T)$ are the current driving direction of V_A and V_B, the "$\cdot$" operation stands for the inner product, and the λ is a predefined factor ($\lambda\in[-1,1]$) used to define the level of divergence. If (2) holds, then V_A selects V_B as a shadow vehicle. From this moment, V_A will send two LBS requests to the LBS server that one with V_A's real location and another with the generated virtual location using the virtual location generation algorithm. Details of the virtual location algorithm will be discussed in the next subsection. If multiple vehicles satisfy the condition, then the V_A will send multiple requests with different location information. V_A will keep using the location information of these shadows to protect its real route until the shadow vehicle gets out of V_A's communication range. At the same time, V_A changes its *PID*. On the other side, V_B does the same procedures as V_A to protect its own privacy. It is a reciprocal process among V_A and its shadow vehicles. Because $Dist(V_A,V_B,T)=Dist(V_B,V_A,T)$ and $\overrightarrow{Dir}(V_B,T)\cdot\overrightarrow{Dir}(V_A,T)=\overrightarrow{Dir}(V_A,T)\cdot\overrightarrow{Dir}(V_B,T)$ are always true, thus the shadow selection process is reciprocal. That means if V_B is selected by V_A as a shadow, then V_A is also selected as a shadow by V_B. When V_A is out of the communication range of V_B, V_B also changes its *PID* simultaneously due to the reciprocal property of the proposed scheme. In this way, the LBS server can hardly determine the real route of either V_A or V_B by linking all the location information extracted from LBS requests. If more vehicles are involved, then the success rate of tracking a vehicle decreases sharply. Algorithm 1 shows the pseudo code of the shadow selection process of the proposed scheme.

C. Virtual Location Generation

After V_A selects its shadow vehicles, V_A needs to generate virtual locations using the traffic conditions of shadow vehicles. Each virtual location will be encapsulated by V_A as a legitimate LBS request and sent to the LBS server. It is desirable that the trace of generated virtual locations diverges from the real trace of the vehicle. In this way, the attacker can hardly guess the real destination of a target vehicle. To achieve this goal, we propose a virtual location generation algorithm as shown in Algorithm 2. The input of Algorithm 2 includes speed and location information of V_A and its shadow vehicle V_i at time T and T^{-1}. The output of the algorithm is the virtual location of V_A generated with the shadow vehicle V_i at time T, which is denoted as $Loc(V_A^i,T)$. The V_A^i represents the virtual vehicle that follows the trace of generated virtual locations.

Algorithm 2: Virtual Location Generation

Input: $Loc(V_A^i, T^{-1}), Loc(V_A, T^{-1}), Loc(V_i, T^{-1}),$
$Loc(V_i, T), SP(V_A^i, T^{-1}), SP(V_A, T^{-1})$
Output: $Loc(V_A^i, T)$
if $Loc(V_A^i, T^{-1}) == NULL$ **then**
  $Loc(V_A^i, T^{-1}) = Loc(V_A, T^{-1})$;
  $SP(V_A^i, T^{-1}) = SP(V_A, T^{-1})$;
end
Calculate $Dist(V_A^i, V_i, T^{-1})$;
if $Dist(V_A^i, V_i, T^{-1}) < Dist_{thld}$ **then**
  $Loc(V_A^i, T) = Loc(V_i, T)$;
else
  $SP(V_A^i, T) = \{SP(V_A, T^{-1}) + acc_{V_A} \cdot \delta,\ SP_{max}\}_{min}$;
  $\overrightarrow{Dir}(V_A^i, T) = Loc(V_i, T^{-1}) - Loc(V_A, T^{-1})$;
  $Loc(V_A^i, T) = Loc(V_A, T^{-1}) + 0.5 \cdot \delta \cdot$
  $[SP(V_A^i, T^{-1}) + SP(V_A^i, T)] \cdot \overrightarrow{Dir}(V_A^i, T)$;
end
return $Loc(V_A^i, T)$

In the algorithm, the first step is to set up the speed and virtual location of V_A^i at T^{-1}. If there is no record, then we set the virtual location of V_A^i the same as the real location of V_A. The initial speed of V_A^i is also set to be the same as V_A. Then, the distance between V_A^i and V_i is calculated and checked if the distance is smaller than a threshold $Dist_{thld}$. The threshold should be a small value such that we can replace the location of V_A^i with V_i without raising suspicion. If the distance is not small enough, then the location of V_A^i at time T is calculated based on the location and speed at T^{-1} as well as the location of V_i. To reduce the distance, we assume V_A^i drives toward the location of V_i using current speed with a constant acceleration acc_{V_A}. The acceleration of the virtual vehicle is set to be the same as the acceleration of V_A. The $\{\cdot\}_{min}$ algorithm is to select the smaller speed value between the two since there is always a speed limit. The new location of V_A^i is calculated by the previous location of V_A^i plus the driving distance within the time slot δ. At last, the algorithm returns the calculated virtual location $Loc(V_A^i, T)$.

V. Performance Analysis

In this section, we conduct simulations to show the effectiveness of the proposed scheme. First, we introduce the privacy metrics that are generally used to evaluate the privacy-preserving schemes. Then, we study the privacy-preserving performance of the proposed scheme with varied vehicle densities. At last, we compare our proposed scheme with other existing schemes.

A. Location Privacy Metrics

In this paper, we use anonymity set size, location entropy, and tracking success ratio to evaluate the level of privacy that can be achieved by the proposed scheme.

1) Anonymity Set Size: After a certain period of time of using the privacy-preserving scheme while having the LBS, V_A has generated many virtual routes with different final locations. The anonymity set of V_A, which is denoted as S_A, is the set of all possible final locations. The size of S_A (denoted as $|S_A|$) is the number of elements in this set. Denote the real location of V_A at time t as $L_A(r,t)$, and denote the ith possible virtual location as $L_A(i,t)$. Let $p_A(i,t)$ be the probability that the attacker regards $L_A(i,t)$ as the final location of V_A, then the anonymity set S_A can be expressed as follows.

$$S_A = \{L_A(r,t), L_A(i,t) | p_A(i,t) \neq 0\}. \tag{3}$$

In this paper, we assume that the attacker's tracking process follows the target tracking algorithms, where the predicted location of time t is mainly determined by the location of the previous time $t-1$[29]. We assume that the attacker has perfect knowledge of the initial location of V_A, thus $p_A(r,0) = 1$. Then, the probability $p_A(i,t)$ that the attacker believes $L_A(i,t)$ is the final location of V_A can be calculated as follows[4].

$$p_A(i,t) = \frac{\sum_{x(\neq i) \in S_A} Dist[L_A(r,t), L_A(x,t)]}{\sum_{y \in S_A} Dist[L_A(r,t), L_A(x,t)]} \times \frac{p_A(j, t-1)}{|S_A| - 1}, \tag{4}$$

where $p_A(j, t-1)$ is the probability associated with the jth location of V_A at the previous time $(t-1)$. This attacker's tracking model uses the distance deviation to make predictions, which ensures $\sum_{i \in S_A} p_A(i,t) = 1$ at any time t.

2) Location Entropy: Location entropy is a precise quantitative metrics to evaluate the location privacy. It represents the degree of uncertainty for the attacker to find the real location $L_A(r,t)$ from S_A. The location privacy is defined as the number of bits that calculated as follows.

$$H_A = -\sum_{i \in S_A} p_A(i,t) \times \log_2(p_A(i,t)). \tag{5}$$

If the location entropy $H_A = n$, that means the attacker thinks that the V_A is equally likely to be in one of the 2^n locations. The larger value of H_A indicates the lower certainty of the attacker or higher location privacy of the vehicle.

3) Tracking Success Ratio: The tracking success ratio describes the probability that the attacker considers $L_A(r,t)$ as the real location of V_A after tracking it over time t. Thus, the tracking success ratio is equivalent to $p_A(r,t)$. Note that the attacker does not know $p_A(r,t)$ since the attacker cannot distinguish the real location from S_A at any time t, except for the initial location $(p_A(r,0) = 1)$.

B. Simulation Results of the Proposed Scheme

Our simulations are based on SUMO[30], OMNet++[31], and Veins[26]. In the simulation settings, 80 to 400 vehicles are randomly distributed in a 6 km $\times$ 6 km region with random

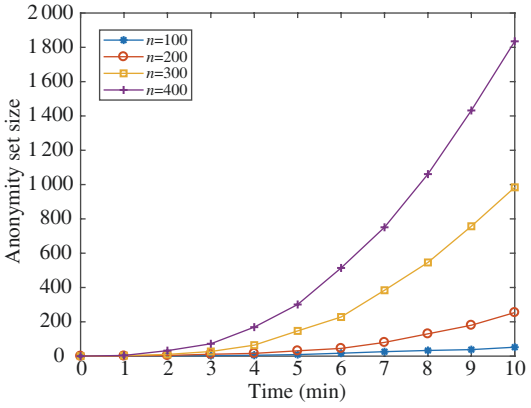

Fig. 4 Anonymity set size in different vehicle densities

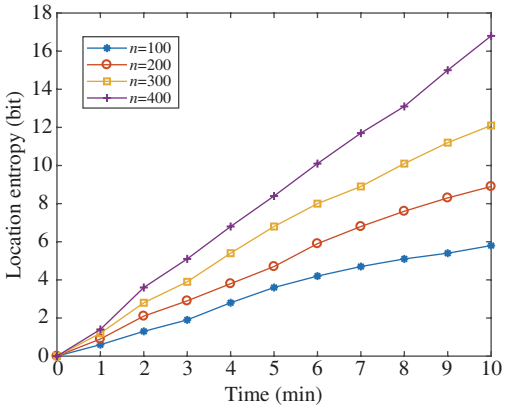

Fig. 5 Location entropy in different vehicle densities

routes. The 6 km × 6 km region is selected from the urban area of Lincoln, Nebraska, USA. All vehicles on the roads send LBS requests and BSMs periodically. The intervals of sending LBS requests and two BSMs are set to be 1 second and 0.2 seconds, respectively. Since the major portion of the simulated region is in a metropolitan area, the vehicle's speed is set to be in the range of 0 to 20 m/s. According to the definition of $Dist_{max}$, we set $Dist_{max} = 20\ m$, which is the maximum distance a vehicle can travel in 1 second with the maximum speed limit. Without loss of generality, the maximum communication range through DSRC, i.e. CR, is set to 300 m, and the $Dist_{thld}$ is set to 5 m. The maximum simulation time is 10 minutes since it is the average trip time from one location to another. The λ value is set to be 0.1 (we consider two vehicles diverge only if the angle between two vehicle's directions is larger than approximately 90 degrees). All the original routes of vehicles are generated by the Randomtrips.py program in SUMO. Other general parameters, like the acceleration of vehicles, follow the default settings in our simulation environment.

Fig. 4. shows the changes of anonymity set size over the time with different vehicle densities. As shown in the figure, when the density of vehicles goes high, the anonymity size grows rapidly as time increases.

Fig. 5. shows the average location entropy of the proposed scheme over time with different vehicle densities. From the figure, we can see that higher vehicle densities achieve a higher level of privacy. When the number of vehicles is 100, after 10 minutes of travel, the location privacy can reach almost 6 bits. According to the definition of the location entropy, in the case of 6 bits entropy, the attacker may find roughly sixty-four locations that are equally like to be the real location of the tracked vehicle. In the scenario with higher vehicle densities, it is much harder for the attacker to find the real location of a vehicle with confidence since the virtual locations are increased exponentially as the increase of location entropy.

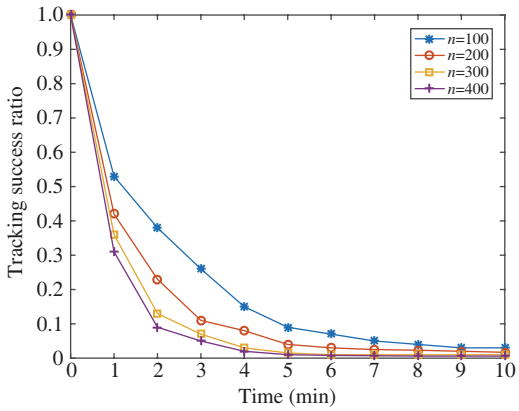

Fig. 6 Tracking success ratio in different vehicle densities

Fig. 6. presents the trend of tracking success ratios over time with different vehicle densities. From the figure, we can see that the tracking success ratio drops rapidly in the first five minutes. In the case with the lowest vehicle density, the tracking success ratio drops below 0.1 after 5 minutes and decreases to approximately 0.02 after 10 minutes. With the higher vehicle density, the tracking success ratio decreases faster and has a lower ratio after 10 minutes of travel. In the case with the highest vehicle density, the tracking success ratio drops below 0.01 before 5 minutes.

In summary, the proposed scheme can achieve strong privacy-preserving results. Even in a low vehicle density scenario, the attacker has a very low tracking success ratio after 10 minutes of travel. Moreover, the proposed scheme is more effective with higher vehicle density. As the number of vehicles increases, the performance of the proposed scheme improves rapidly. Because once the total number of vehicles is increased, the vehicle density is increased. Then, each vehicle will have more surrounding vehicles during the trip. As the number of surrounding vehicles increases, the number of shadow vehicles that can be selected to cover the real route is increased as well. Although we cannot directly control the number of surrounding vehicles to find its relationship with

the performance gain, the simulation results indicate that the performance gain caused by the increasing in vehicle number is close to exponential growth.

C. Performance Comparison

We also compare our proposed scheme with the MOP scheme and the PPRLD scheme from the aspects of location entropy and tracking success ratio. The simulation results of 10 minutes traffic are shown in. Fig. 7 and Fig. 8. with the number of vehicles set to 240. The Fig. 9 and Fig. 10 compare the performance of the proposed scheme and other schemes under different number of vehicles after 10 minutes simulation.

From Fig. 7 we can see that the proposed scheme has higher location entropy than the MOP scheme and the PPRLD scheme during the whole trip. After 10 minutes, the location entropy of the proposed scheme is around 2 bits higher than the PPRLD scheme, and around 4.5 bits higher than the MOP scheme. From Fig. 8 we can see that for all three schemes, the tracking success ratio drops rapidly at the first 5 minutes, and decreases to around 0.01 after 10 minutes. However, if we observe the results horizontally, we can observe that the proposed scheme can reach the same tracking success ratio around one minute earlier than the PPRLD scheme and 2 minutes earlier than the MOP scheme.

In Fig. 9, the location entropy is increased almost linearly as the number of vehicles increases. By the definition of location entropy, the number of plausible locations of a specific vehicle increases exponentially as the number of vehicles increases. Simulation results also show that the proposed scheme can always achieve higher location entropy compared with the PPRLD and the MOP scheme. In Fig. 10, we compare the tracking success ratio under different number of vehicles. All three schemes can achieve a low tracking success ratio when the number of vehicles is high after 10 minutes simulation. With the same number of vehicles in simulation, the proposed scheme can always achieve a lower tracking success ratio than the other two schemes.

Overall speaking, our proposed scheme can achieve a higher privacy level than the MOP scheme and the PPRLD scheme. This is because that the proposed scheme allows multiple vehicles to act as shadows at one time instead of only one shadow that is used in the MOP and the PPRLD scheme. Moreover, the proposed scheme does not require establishing an agreement between the shadow and the tracking vehicle, which makes it more flexible and efficient. The MOP scheme is mainly based on the prediction of the future route of the candidate shadow, and the generated virtual route is also determined by prediction, which may be vulnerable to movement tracking attacks. The proposed scheme uses the real location of another vehicle as its virtual location, which is more convincing. In the PPRLD scheme, the shadow vehicle

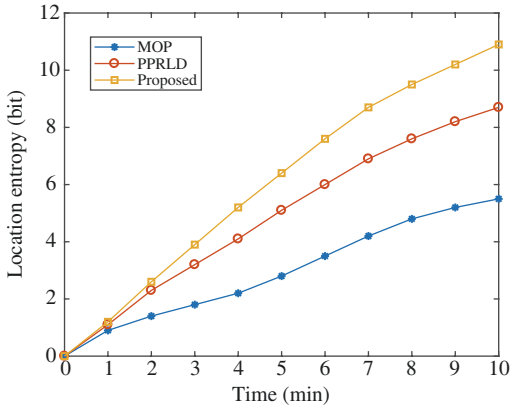

Fig. 7 Comparison of location entropy

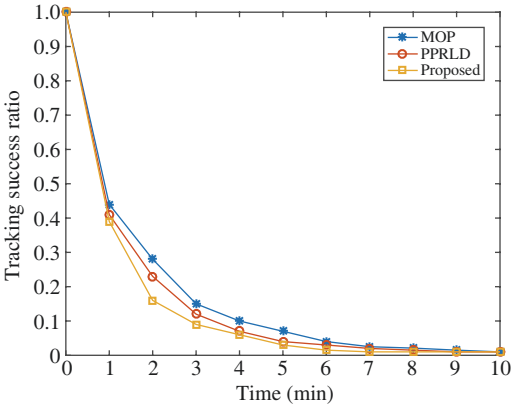

Fig. 8 Comparison of tracking success ratio

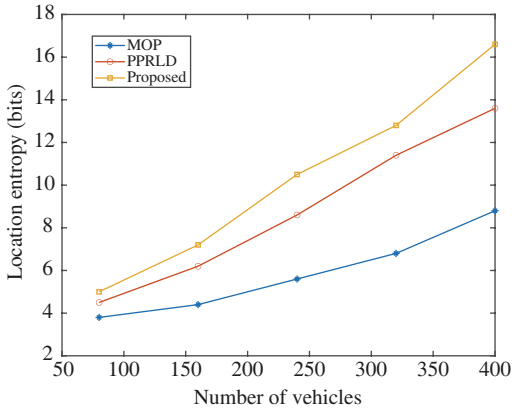

Fig. 9 Location entropy under different vehicle density

will converge to the real route of the tracking vehicle. It is highly possible for the attacker to analyze all possible routes and identify the one with the longest length to be the real route since all generated virtual routes would eventually merge into the real route. In contrast to that, the virtual routes generated in our proposed scheme diverge from the real route, making it harder for the attacker to track a vehicle.

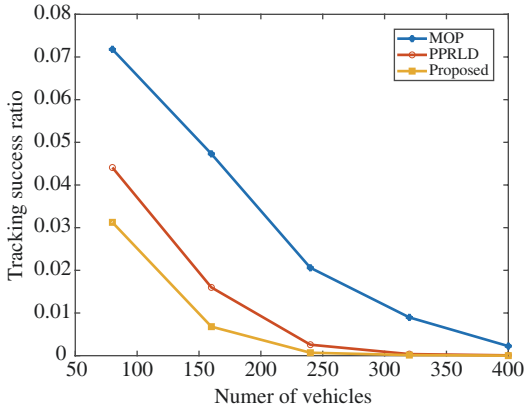

Fig. 10 Tracking success ratio under different vehicle density

VI. Conclusion

With the flourish of various LBSs provided in IoV, the large amount of user location information that uploaded to the LBS server has put user privacy at risk. Thus, protecting user location information from being exposed to attackers has been an open topic for a long time. In this paper, we summarized and compared different kinds of location privacy schemes. Although many solutions have been proposed in the literature, only a few of them can preserve location privacy while having real-time LBS with a frequent and accurate location information update. Observing the limitation of the existing schemes, we proposed a novel scheme which not only allows user to have real-time LBS with accurate location update but also achieves higher privacy level than existing schemes. Simulation results show that our proposed scheme can provide satisfactory location privacy protection when vehicle density is low, and the performance becomes much better in the scenarios with high vehicle densities. Simulation results show that our proposed scheme outperforms other schemes in terms of location entropy and tracking success ratio.

References

[1] CONTRERAS-CASTILLO J, ZEADALLY S, GUERRERO-IBAÑEZ J A. Internet of vehicles: architecture, protocols, and security[J]. IEEE Internet of Things Journal, 2018, 5(5): 3701-3709.

[2] ZHENG Y L, LUO J, ZHONG T. Service recommendation middleware based on location privacy protection in VANET[J]. IEEE Access, 2020, 8: 12768-12783.

[3] HUANG J Q, FANG D F, QIAN Y, et al. Recent advances and challenges in security and privacy for V2X communications[J]. Open Journal of Vehicular Technology, 2020, 1(1): 244-266.

[4] LIM J, YU H, KIM K, et al. Preserving location privacy of connected vehicles with highly accurate location updates[J]. IEEE Communications Letters, 2017, 21(3): 540-543.

[5] SWEENEY L. K-anonymity: A model for protecting privacy[J]. International Journal of Uncertainty, Fuzziness and Knowledge Based System, 2002, 10(5): 557-570.

[6] GEDIK B, LIU L. Protecting location privacy with personalized k-anonymity: architecture and algorithms[J]. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18.

[7] NIU B, LI Q H, ZHU X Y, et al. Achieving k-anonymity in privacy-aware location-based services[C]//IEEE INFOCOM 2014, 2014: 754-762.

[8] FÖRSTER D, LÖHR H, KARGL F. Decentralized enforcement of k-anonymity for location privacy using secret sharing[C]//2015 IEEE Vehicular Networking Conference (VNC), 2015: 279-286.

[9] WANG J B, CAI Z P, YU J G. Achieving personalized k-anonymity-based content privacy for autonomous vehicles in CPS[J]. IEEE Transactions on Industrial Informatics, 2020, 16(6): 4242-4251.

[10] GEDIK B, LIU L. Location privacy in mobile systems: a personalized anonymization model[C]//25th IEEE International Conference on Distributed Computing Systems (ICDCS'05), 2005: 620-629.

[11] HUANG L P, MATSUURA K, YAMANE H, et al. Enhancing wireless location privacy using silent period[C]//IEEE Wireless Communications and Networking Conference, 2005: 1187-1192.

[12] SAMPIGETHAYA K, LI M Y, HUANG L P, et al. AMOEBA: robust location privacy scheme for VANET[J]. IEEE Journal on Selected Areas in Communications, 2007, 25(8): 1569-1589.

[13] BUTTYÁN L, HOLCZER T, WEIMERSKIRCH A, et al. SLOW: a practical pseudonym changing scheme for location privacy in VANETs[C]//2009 IEEE Vehicular Networking Conference (VNC), 2009: 1-8.

[14] BERESFORD A R, STAJANO F. Location privacy in pervasive computing[J]. IEEE Pervasive Computing, 2003, 2(1): 46-55.

[15] CARIANHA A M, BARRETO L P, LIMA G. Improving location privacy in mix-zones for VANETs[C]//30th IEEE International Performance Computing and Communications Conference, 2011: 1-6.

[16] LU R X, LIN X D, LUAN T H, et al. Pseudonym changing at social spots: an effective strategy for location privacy in VANETs[J]. IEEE Transactions on Vehicular Technology, 2012, 61(1): 86-96.

[17] GUO N, MA L Y, GAO T H. Independent mix zone for location privacy in vehicular networks[J]. IEEE Access, 2018, 6: 16842-16850.

[18] ULLAH I, WAHID A, SHAHM A, et al. VBPC: velocity based pseudonym changing strategy to protect location privacy of vehicles in VANET[C]//2017 International Conference on Communication Technologies (ComTech), 2017: 132-137.

[19] NIU B, LI Q H, ZHU X Y, et al. Enhancing privacy through caching in location-based services[C]//2015 IEEE Conference on Computer Communications (INFOCOM), 2015: 1017-1025.

[20] LIU B, ZHOU W L, ZHU T Q, et al. Silence is golden: enhancing privacy of location-based services by content broadcasting and active caching in wireless vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(12): 9942-9953.

[21] FANG S S, MAO H N. A Connectivity-aware caching algorithm for vehicular content centric networks with cache-enabled vehicles[C]//2018 IEEE/CIC International Conference on Communications in China (ICCC Workshops), 2018: 232-236.

[22] 5.9 GHz DSRC connected vehicles for intelligent transportation systems. https://ecfsapi.fcc.gov/file/7520943378.pdf

[23] CUI J, WEN J Y, HAN S S, et al. Efficient privacy-preserving scheme for real-time location data in vehicular ad-hoc network[J]. IEEE Internet of Things Journal, 2018, 5(5): 3491-3498.

[24] BABAGHAYOU M, LABRAOUI N, ARI A A A, et al. The impact of the adversary's eavesdropping stations on the location privacy level in Internet of vehicles[C]//2020 5th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), 2020: 1-6.

[25] EMARA K. Poster: PREXT: Privacy extension for Veins VANET simulator[C]//2016 IEEE Vehicular Networking Conference (VNC), 2016:

1-2.
[26] SOMMER C, GERMAN R, DRESSLER F. Bidirectionally coupled network and road traffic simulation for improved IVC analysis[J]. IEEE 1030 Trans. Mobile Comput., 2011, 10(1): 3C15.
[27] WANG F, XU Y J, ZHANG H W, et al. 2FLIP: A two-factor lightweight privacy-preserving authentication scheme for VANET[J]. IEEE Transactions on Vehicular Technology, 2016, 65(2): 896-911.
[28] HUANG J Q, QIAN Y, HU Q Y. Secure and efficient privacy-preserving authentication scheme for 5G software defined vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2020, 69(8): 8542-8554.
[29] HOH B, GRUTESER M. Preserving privacy in GPS traces via uncertainty-aware path cloaking[C]//Proceedings of the 14th ACM conference on Computer and communications security, 2007:161C171.
[30] LOPEZ P A, BEHRISCH M, BIEKER-WALZ L, et al. Microscopic traffic simulation using SUMO[C]//Proc. 1033 IEEE Intell. Transp. Syst. Conf., 2018: 2575C2582.
[31] VARGA A. The OMNeT++ discrete event simulation system[C]//European Simulation Multiconference (ESM'2001), 2001.

About the Authors

Jiaqi Huang received his B.S. degree in Spatial Informatics & Digitalized Technology from the University of Electronic Science and Technology of China in 2016. He received his Ph.D. degree in Computer Engineering from the University of Nebraska-Lincoln in 2020. He joined the School of Computer Science and Mathematics as an assistant professor in 2020. His research interests include cybersecurity and network security, vehicular networks, 5G networks, fog computing, and Internet of Things.

Yi Qian received a Ph.D. degree in electrical engineering from Clemson University. He is a professor in the Department of Electrical and Computer Engineering, University of Nebraska-Lincoln (UNL). Prior to joining UNL, he worked in the telecommunications industry, academia, and the government. Some of his previous professional positions include serving as a senior member of scientific staff and a technical advisor at Nortel Networks, a senior systems engineer and a technical advisor at several start-up companies, an assistant professor at University of Puerto Rico at Mayaguez, and a senior researcher at National Institute of Standards and Technology. His research interests include information assurance and network security, network design, network modeling, simulation and performance analysis for next generation wireless networks, wireless adhoc and sensor networks, vehicular networks, smart grid communication networks, broadband satellite networks, optical networks, high-speed networks and the Internet. Prof. Yi Qian was the Chair of IEEE Communications Society Technical Committee for Communications and Information Security from January 1, 2014 to December 31, 2015. He was the Technical Program Chair for IEEE International Conference on Communications (ICC) 2018. He is serving on the editorial boards for several international journals and magazines, including serving as the Editor-in-Chief for IEEE Wireless Communications Magazine. He was a Distinguished Lecturer for IEEE Vehicular Technology Society. He is currently a Distinguished Lecturer for IEEE Communications Society.

Rose Qingyang Hu is a Professor in the Electrical and Computer Engineering Department and Associate Dean for research of College of Engineering at Utah State University. She also directs Communications Network Innovation Lab at Utah State University. Her current research interests include next-generation wireless system design and optimization, Internet of Things, cyber physical system, mobile edge computing, V2X communications, artificial intelligence in wireless networks, wireless system modeling and performance analysis. Prof. Hu received the B.S. degree from the University of Science and Technology of China, the M.S. degree from New York University, and the Ph.D. degree from the University of Kansas. Besides a decade academia experience, she has more than 10 years of R&D experience with Nortel, Blackberry, and Intel as a Technical Manager, a Senior Wireless System Architect, and a Senior Research Scientist, actively participating in industrial 3G/4G technology development, standardization, system level simulation, and performance evaluation. She has published extensively in top IEEE journals and conferences and also holds numerous patents in her research areas. Prof. Hu is currently serving on the editorial boards of the IEEE Transactions on Wireless Communications, the IEEE Transactions on Vehicular Technology, the IEEE Communications Magazine and the IEEE Wireless Communications. She also served as the TPC Co-Chair for the IEEE ICC 2018. She is an IEEE Communications Society Distinguished Lecturer Class 2015-2018 and a recipient of prestigious Best Paper Awards from the IEEE GLOBECOM 2012, the IEEE ICC 2015, the IEEE VTC Spring 2016, and the IEEE ICC 2016. Prof. Hu is a Fellow of IEEE and a member of Phi Kappa Phi Honor Society.