Location and trajectory privacy preservation in 5G-Enabled vehicle social network services

Dan Liao, Hui Li, Gang Sun, Ming Zhang, Victor Chang

PII: S1084-8045(18)30039-0
DOI: 10.1016/j.jnca.2018.02.002
Reference: YJNCA 2061

To appear in: Journal of Network and Computer Applications

Received Date: 10 April 2017
Revised Date: 16 January 2018
Accepted Date: 2 February 2018


This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.
Location and Trajectory Privacy Preservation in 5G-Enabled Vehicle Social Network Services

Dan Liao¹, Hui Li¹,², Gang Sun¹,³, Ming Zhang², Victor Chang⁴

¹Key Lab of Optical Fiber Sensing and Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu, China
²Chengdu Research Institute of UESTC, Chengdu, China
³Center for Cyber Security, University of Electronic Science and Technology of China, Chengdu, China
⁴Xi’an Jiaotong Liverpool University, Suzhou, China

Abstract: 5G-based Vehicular Social Networks (VSNs) demand an advanced location and trajectory privacy preserving scheme for vehicles. Because VSNs present the characteristics of high mobility and multiple hop relays, we design a 5G-based VSN framework that incorporates Mobile Femtocell (MFemtocell) technology. Then, we propose the Dynamic Group Division algorithm (DGD), which is suitable for the dynamic properties of 5G and meets the real-time demands of VSN. To preserve privacy, the DGD algorithm increases the likelihood of exchanging pseudonyms via the proposed Group Generating Protocol and Pseudonym Exchanging Protocol. Then, we adopt the composite metric KDT (where K is the average anonymity set size, D is the average distance deviation, and T is the anonymity duration) and pseudonym entropy to quantify the degree of privacy. We evaluate and validate the effectiveness of our proposed algorithm based on the following three aspects: anonymity set size, distance deviation and pseudonym entropy. The simulation results show that our DGD algorithm better protects the location and trajectory privacy of VSNs while sustaining higher real-time demand than current approaches.

Key words: Location privacy; Trajectory privacy; 5G; Mobile Femtocell; VSN

1. INTRODUCTION

Vehicles now represent “the biggest mobile terminal” in the context of the Internet. The 5GAA committee (5G Automotive Alliance) has published research on the Internet of Cars in 2017 [1], and with the development of cloud computing [2-7] and big data [8-13], vehicles represent intelligent devices that can connect to the Internet and present sensing and computing abilities. Thus, vehicles are now the main carriers for mobile social networks. The Vehicular Social Network (VSN) has emerged, and vehicles can now be connected to wireless networks to improve traffic safety and promote the development of smart cars. However, the convenience of VSNs may lead to privacy concerns. The problem of privacy disclosure primarily stems from two aspects. First, for the users, certain data transmitted over the VSN are highly sensitive, such as location, trajectory, and identity information. If these sensitive data are revealed, the location privacy, trajectory privacy and identity privacy can be leaked [14]. Second, the topology of the VSN changes quickly because of the vehicle's high-speed mobility. Thus, data transmission exploits the multiple hop relay method. However, multiple hop relays are prone to data leakage risks, which may lead to the leakage of private information. Furthermore, people are paying increasing attention to their own privacy and data security [15-21]. Therefore, this paper addresses the problem of privacy leakage in VSNs. Combined with modern communication technology (5G), the method in this paper effectively protects the vehicles’ location and trajectory privacy in the VSN.

With the increasing number of connected devices and demand for data rate, the 5G wireless communication system has been a popular research area in recent years [22-23]. The next 5G can serve all types of applications/systems with extremely high user rates anytime and anywhere [24]. As a Wireless Sensor Network (WSN), VSNs will inevitably lead to extraordinary developments with the application of 5G. Compared with other WSNs, VSNs realize the modern Intelligent Transport System (ITS). However, VSNs have inherent characteristics, such as high mobility and multiple hop relays. Thus, Mobile Femtocell (MFemtocell) has been introduced for 5G technology [25]. The use of MFemtocell can significantly maximize performance, such as by realizing dynamic linking, enhancing user throughput, and reducing response times and signal overhead [26-27].

To reduce traffic accidents, vehicles send safety message periodically. The safety message includes information about the location, speed and direction of the vehicles. Although the VSN can be plugged into the 5G network seamlessly, the 5G-based VSN does not consider privacy preservation. If a malicious attacker continuously eavesdrops on the safety message, the location and trajectory privacy may be leaked. To address this problem, researchers have proposed efficient schemes that include K-anonymity [28], Mix-zone [29], MixGroup [30], and Encryption [31-32]. The basic ideas behind these schemes are consistent. Each vehicle is assigned a pseudonym in the VSN, and then vehicles exchange the
By combining the Mix-zone and group signature technology, the MixGroup algorithm extends the group region and increases the opportunities for exchanging pseudonyms to protect location privacy as shown in Fig. 1. Nevertheless, the division of the group region in the MixGroup method has three main shortcomings. First, the response time of MixGroup cannot meet the real-time demand of the VSN. Second, the static division of the group region (as shown in Fig. 1) cannot be applied perfectly in future 5G architecture, because 5G MFemtocell technology is dynamic and allows the vehicle to adaptively access the core network. Third, because the division of the group region is based on the vehicles’ routes while moving, the MixGroup method cannot effectively protect the vehicles’ trajectory privacy. User behavior is known to have certain regular characteristics over a long period of time. Because of work or family circumstances, vehicles regularly drive past a number of fixed locations or traverse along fixed trajectories. Because of this certain regularity, the division of the group region is vulnerable to leak trajectory privacy in the MixGroup method.

Fig. 1: MixGroup method

To address the main shortcomings of the MixGroup method, we design a novel 5G-based VSN architecture and propose the Dynamic Group Division algorithm (DGD) for the protection of vehicle location and trajectory privacy. The main contributions of this paper are as follows.

- To improve the real-time demand of the VSN, we combine 5G wireless communication systems and VSNs. This paper includes MFemtocell technology and designs a novel system model, the 5G-based VSNs.
- To quantify the location and trajectory privacy that can be suitably applied in the 5G-based VSNs, this paper improves the composite metric KDT, where $K$ is the average anonymity set size, $D$ is the average distance deviation, and $T$ is the anonymity duration.
- To more effectively protect privacy, we propose the DGD algorithm for the 5G-based VSNs. Based on social/individual hot spots, the DGD algorithm designs the Group Generating Protocol and the Pseudonym Exchanging Protocol to dynamically divide the group region and increase the probability of exchanging pseudonyms.

The remainder of this paper is organized as follows. In Section 2, we review related works on the development of integration between 5G networks and VSNs and schemes for preserving privacy in VSNs. In Section 3, we describe the basic concepts and relative definitions. In Section 4, the motivation, objective and system model are revealed. In Section 5, a detailed description of the DGD algorithm is provided. In Section 6, the simulation environment and results are presented. In Section 7, the conclusions of this paper are provided.

2. RELATED WORK

Significant attention has been focused on the problems of privacy preservation and real-time demand in VSNs [33-35]. With the constant development of communication technology, higher speeds and greater bandwidth have promoted the development of integration between the 5G network and VSNs. Through IEEE 802.11p, the work in [36] proposes a 5G-VANET (Vehicular Ad Hoc Network) architecture to realize adaptive clustering, which can reduce the system response time and improve the real-time demand of VANET. The authors in [37] address the security and privacy issues in 5G-enabled vehicular networks and propose the scheme of a real-time video reporting service in a 5G network framework. For protecting the privacy of a video reporting service in vehicular networks, the authors exploit the method of pseudonym management combined with the high speed and low delay characteristics of the 5G network. The study presented in [38] presents a 5G-enabled vehicular network and introduces 5G technologies, such as Device-to-Device (D2D), Enhanced Cloud Radio Access Network (EC-RAN) and Heterogeneous Network (HetNet), into the vehicular network. Moreover, C. Wang, et al. [25] propose a MFemtocell network. MFemtocell is a 5G technology that combines mobile relays with Femtocells. MFemtocell can be deployed in vehicular networks to provide enhanced throughput and spectrum efficiency. With the integrated application of the 5G and vehicular network, MFemtocell allows for the demands of VSNs to be met in real-time and with heavy data traffic.

In VSNs, vehicles can easily access location-based services for the convenience of drivers [39-40], such as finding the nearest gas station or supermarket. Therefore, protecting the location and trajectory privacy of vehicles has received increasing attention recently. With the technology of homomorphic encryption, the work in [41] realizes privacy preservation through the sharing of encryption and trajectories between vehicles. However, because of this sharing, the approach [41] works on the condition that the density of vehicles is high. Other approaches [42-48] have adopted the authentication mechanism to meet the requirement of privacy protection. Facing the challenges of real-time demand in VSNs, researchers have also proposed...
lightweight authentication mechanisms. For example, Wang [44] proposes a Two-Factor Lightweight Privacy-preserving authentication scheme (2FLIP) that employs the decentralized CA and biological-password based two-factor authentication. Although the 2FLIP scheme can reduce the costs of authentication and realize conditional privacy protection, the safety of the whole scheme heavily relies on the unique system key of the CA. Abbas [46] proposed a hierarchical pseudonymous authentication protocol that divides pseudonyms into two sub-ranges, primary and secondary, based on the time. The lifecycle of the primary pseudonym is longer than that of the secondary pseudonym. The primary pseudonym is used to communicate with partially trusted institutions, whereas the secondary pseudonym is used between vehicles. This scheme is beneficial to reduce the system burden of the VSN, and it is no longer dependent solely on the CA.

Although the schemes mentioned above have addressed preserving the privacy of the VSN under certain conditions, they are inefficient when the VSN is applied in 5G networks. Moreover, most studies have barely focused on location privacy preservation and have ignored the trajectory privacy of moving vehicles. Under the synthetic consideration of real-time demands and privacy preservation, this paper addresses the issue of location and trajectory privacy in 5G-based VSNs.

3. Preliminaries

In this section, we provide a detailed introduction to certain basic concepts relevant to MFemtocell technology and social/individual hot spots. The definition of KDT used in this paper is subsequently provided.

A. Mobile Femtocell

To meet the high-mobility demand of users, the architecture of the 5G network incorporates the technology of MFemtocell, which combines the mobile relay and femtocell network. The wireless interface of MFemtocell is backwards compatible with all the existing terminal equipment. Thus, MFemtocell can be seamlessly connected to VSNs. As shown in Fig. 2, MFemtocell is always deployed on fast-moving vehicles, such as cars, trains, buses, etc. With the help of MFemtocell, vehicles can adaptively communicate with other vehicles or the Base Station. Therefore, this technology can improve the throughput and reduce the system response rate. Furthermore, MFemtocell can dynamically access the wireless network and effectively resolve the problems caused by the topology of VSNs, which do not have a center node and change quickly.

B. Social hot spots and individual hot spots

A hot spot is a location that users are interested in. In this paper, all locations can be classified into two categories: social hot spot and individual hot spot. Social hot spots are characterized by a location with a high density of vehicles, such as crossroads, parking lots, etc., where vehicles easily encounter each other. Because many meeting opportunities occur for vehicles in social hot spots, existing privacy protection methods (e.g., Mix-Zone) usually take advantage of social hot spots for exchanging the pseudonyms of vehicles. Then, malicious attackers cannot guess the vehicle’s real identity, and the vehicles’ location and trajectory privacy are protected.

Although individual hot spots are characterized by frequent vehicle visits, most people generally drive to and from their places of work and their homes over a long period of time. Therefore, most drivers regularly visit certain fixed locations, such as the supermarket near the driver’s home or the gas station on the way to work as shown in Fig. 3. Of course, one location may be an individual hot spot as well as a social hot spot for certain drivers, and when this occurs, we define the location as a social hot spot by default.

In this paper, the location of a hot spot (including social hot spot and individual hot spot) is denoted by \( PL(x, y) \), where \( x \) and \( y \) represent the latitude and longitude of the area of the hot spot, respectively. To effectively protect vehicle location and trajectory privacy in the 5G-based VSNs, our paper dynamically divides the group area by the definition of \( PL \).

C. Quantification for trajectory privacy

Trajectory privacy is the degree to which an entity cannot be linked to the vehicles’ trajectory over a consecutive time. In this paper, we adopt a composite metric \( KDT \) [49] to
quantify the trajectory privacy, where $K$ refers to the average anonymity set size, $D$ is the average distance deviation, and $T$ is the anonymity duration. Considering realistic demands, we redefine KDT as follows.

1) Anonymity Duration $T$: In our paper, for a better description of the trajectory privacy, we define $T$ as a constant. $T$ is a set of time $\{t_0, t_1, ..., t_n\}$. Therefore, a vehicle is driven for a continuous time $T$ (starting from $t_0$ to $t_n$), and generates a trajectory.

2) Average Anonymity Set Size $K$: Assume that $K_i$ represents the number of vehicles in a group area at time $t_i$. Then, we can find the expression of average anonymity set size $K$.

$$K = \frac{K_0 + K_1 + ... + K_i + ... + K_n}{n+1}, t_i > 0, i = 0, 1, 2, ..., n \quad (1)$$

According to the knowledge of random processes, we know that the number of vehicles that arrive at the group region follows a Poisson distribution $\{N(t)\}$. The one-dimensional probability distribution of $N(t)$ is as follows:

$$P\{ N(t) = k \} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}, k = K_k \cdots K_i \quad (2)$$

Therefore, we can calculate the expectation $E[N(t)] = \lambda t$, where $t \in \{t_0, t_1, ..., t_n\}$, and $\lambda$ is the average number of vehicles arriving in the group area per unit time.

3) Average Distance Deviation $D$: At time $t_i$, suppose that there are two vehicles $(u_m, u_t)$ in a group area and the number of vehicles in the group is $K_i$. We define the distance $d_{mn}$ between two vehicles $(u_m, u_t)$:

$$d_{mn} = \sqrt{(x_m - x_t)^2 + (y_m - y_t)^2} \quad (3)$$

where $(x_m, y_m)$ is the location of vehicle $u_m$ and $(x_t, y_t)$ is the location of vehicle $u_t$. Assume that $P_{mn}$ is the probability of identity confusion between vehicles. That is, an attacker guesses that vehicle $u_m$ is vehicle $u_t$ with probability $P_{mn}$.

Thus, we can obtain the distance deviation $d_{mn}$ at time $t_i$:

$$d_{mn} = \frac{1}{K_i} \sum_{m=1}^{K_i} \sum_{k=1}^{K_i} d_{mn} P_{mn} \quad (4)$$

Then, the average distance deviation $D$ is calculated at the anonymity duration $T$:

$$D = \frac{d_{1} + d_{2} + ... + d_{m} + ... + d_{n}}{n+1}, t_i > 0, i = 0, 1, 2, ..., n \quad (5)$$

For the best trajectory preservation, we must maximize the average anonymity set size $K$ and the average distance deviation $D$ for a continuous time $T$ (Anonymity Duration). Greater average anonymity set sizes and average distance deviations correspond to better operability of anonymous schemes for vehicles. Therefore, this scheme can more effectively protect the trajectory privacy. However, $K$ and $D$ are not infinite values. Therefore, assume that the upper limit of $K_t$ and $d_i$ are $K_{max}$ and $d_{max}$, respectively. The values of $K_{max}$ and $d_{max}$ are substituted into formula (1) and (5). In addition, the maximum of $K$ and $D$ are also $K_{max}$ and $d_{max}$, respectively.

4. MOTIVATION AND SYSTEM MODEL

This section provides the motivation and objective of this paper. Then, we design a 5G-based VSN system model for solving the problem of real-time demand and privacy preservation.

4.1 Motivation

The construction of the MixGroup framework extends the pseudonym-changing regions with the Mix-zone and group signature mechanism, and it addresses the location privacy issue in VSNs. However, the MixGroup framework does not consider a method of protecting trajectory privacy. According to the vehicles’ paths while moving, MixGroup statically generates a group region. Assume that a malicious attacker has joined in the group region. Then, he/she will constantly exchange pseudonyms with other vehicles located in the same group region and will keep trying to exchange pseudonyms until leaving the group region. The malicious attacker can draw the outline of the group region. By matching group regions with the map (e.g., Google Maps), the attacker learns the trajectories of vehicles. If the attacker obtains other information, such as the work location, personal interests, etc., he/she can easily know the real trajectory of the vehicles. There is no guarantee that the users’ trajectories will not be exposed; thus, the users’ trajectory privacy is leaked. Therefore, preserving privacy is insufficient in the MixGroup framework. In this paper, we aim at protecting the location and trajectory privacy in VSNs.

Furthermore, the real-time demand of the VSN is significant. Therefore, using a VSN does not make sense unless the vehicles obtain a quality-of-service guarantee for real-time services. To improve the pseudonym exchange opportunities, the MixGroup framework generates group regions with the static methods. The MixGroup framework does not consider the time cost. However, because of the development of mobile communication technology, an increasing number of vehicles have accessed VSNs. Therefore, the static division of the group region is not appropriate for the rapid development of the Internet. To support the large amount of data traffic and meet real-time demands, this paper introduces 5G MFemtocell technology. MFemtocell can dynamically connect to an operator’s core network. This dynamic characteristic will greatly improve the system response time and meet the real-time demand in VSNs.

4.2 Objective

In this paper, we aim to propose the DGD algorithm for
protecting location and trajectory privacy in 5G-based VSNs. First, we build a novel system model with the introduction of the MFemtocell. Then, with the help of MFemtocell, the vehicles can be dynamically brought together for generating a group in time $T$ (Anonymity Duration). The average anonymity set size $K$ and the average distance deviation $D$ of the group must meet certain conditions for reaching the degree of privacy required by the users. Finally, the DGD algorithm exchanges the identities in the same group with pseudonym entropy principle. By increasing the opportunities to exchange vehicle identities, the ability of an attacker to determine the real identity of a vehicle is hindered.

Thus, based on the MFemtocell technology, hot spot concept and composite metric KDT, the problem of protecting a vehicles’ location and trajectory privacy is resolved by maximizing the average anonymity set size $K$ and the average distance deviation $D$ for a continuous time $T$ (Anonymity Duration) in the 5G-based VSNs.

### 4.3 System model

To meet the real-time demand of vehicles and preserve vehicular privacy, we introduce the MFemtocell technology and design a VSN model, and its framework is shown in Fig. 4. The framework consists of three key components: Vehicles, Registration Authority (RA) and Base Station.

![Fig. 4: Framework of the 5G-based VSNs](image)

Vehicles: the moving entities in the VSN. Each vehicle is equipped with a MFemtocell for communication between vehicles or with the users inside the vehicle.

RA: In this paper, the RA is a strictly trusted third party entity, and it can potentially benefit the system security by generating secure parameters [$PID$, $PK/SK$, $Cert$] for the entities in the proposed system model, where $PID$ is the pseudonym identity for the entities, such as vehicles, groups, etc.; $PK/SK$ is the public-private key pair used to encrypt or decrypt messages to ensure information security; and $Cert$ is the certification parameter applied in the group signature mechanism for preventing the information from being tampered with, forged or imitated.

Base Station: located between the vehicular network and the core network. The Base Station can directly communicate with the MFemtocell for collecting information from vehicles. Then, the Base Station sends the gathered messages to the core network where they are processed, and then they are fed back to the Base Station.

Fig. 4 shows that the framework of the 5G-based VSN has been divided into four main stages: system initialization, group generating, pseudonym exchange, and group cancellation.

1) **System initialization**: In our proposed system model, the RA is a fully trusted entity and it boots and initializes the whole system. The RA can verify the legitimacy of each entity in the VSN and assign the corresponding secure parameters $PID$, $PK/SK$, $Cert$ for the legal entities. Of course, the RA server stores the vehicles’ true identities and pseudonym identities, which contributes to tracking offenders for law enforcement. Because the RA is a trusted entity, the problem of privacy leakage from the RA server is not an issue.

2) **Group generating**: If a moving vehicle $v_i$ sends the safety message periodically, MFemtocell technology allows the vehicle $v_i$ to dynamically generate group regions along the vehicle’s trajectory. If the vehicle $v_i$ sends a safety message in one location, then the group region is generated based on the values of $K_{\text{max}}$ and $d_{\text{max}}$.

3) **Pseudonym exchange**: In the group region, vehicles communicate with a group identity between the Base Station and vehicles. To protect the location and trajectory privacy, vehicles located in the same group region need to exchange pseudonyms with each other constantly. Thus, vehicles will obtain a different pseudonym after moving out of a group region every time, which increases the difficulty for a malicious attacker to guess the real identity of the vehicles.

4) **Group cancellation**: With the moving of the vehicle, the group region will be changed constantly along the trajectory of the vehicle. Therefore, the group region is generated and revoked alternately. When a group cannot meet the requirements of the system (e.g., anonymity entropy), the group will be revoked automatically with the help of MFemtocell and the source of the group will be recycled by the RA.

### 5. Algorithm Design

In this section, we first present the framework of the DGD algorithm. Then, we introduce three protocols in the DGD algorithm accordingly: Group Generating Protocol, Pseudonym Exchanging Protocol and Group Revocation Protocol.

#### 5.1 DGD algorithm

Fig. 5 describes the framework of the DGD algorithm. The DGD algorithm addresses the location and trajectory privacy in the 5G-based VSNs. When a vehicle sends the safety message periodically from one location, the location privacy is protected. First, the vehicle generates a group region using the method that is introduced in the Group Generating...
Protocol. Then, to protect the location privacy, all vehicles exchange pseudonyms with each other in the same group. When a vehicle is moving, location privacy as well as trajectory privacy should be considered. To dynamically generate group regions along the trajectory, the DGD algorithm calls the Group Generating Protocol and Pseudonym Exchanging Protocol alternately.

Algorithm 1 describes the pseudo code of the DGD algorithm.

Algorithm 1: Dynamic Group Division (DGD)

1: if (vehicle is moving)
2:   Call for the Group Generating Protocol alternately;
3:   Dynamically obtain group regions;
4:   if (One group region is unsuccessful)
5:      Go to the scheme of location privacy;
6:   else
7:    Call for Pseudonym Exchanging Protocol alternately;
8:   end if
9: else
10:    Generate a group region for the vehicle;
11:    Exchange pseudonyms with each other;
12: end if
13: Call for Group Revocation Protocol.

Algorithm 2 describes the pseudo code of the Group Generating Protocol in details.

Algorithm 2: Group Generating Protocol

Input: PID_v, PK_v, SK_v, Cert_v, PL, t_stamp, K_max, d_max
1: A vehicle v_i enters a hot spot and listens for the request message for cell generating;
2: for (t_stamp >= 0)
3:   if (vehicle v_i receives the message Cell_Generate_Request)
4:      Join in the cell;
5:   else
6:      Initiate the request message Cell_Generate_Request=PID_v,

Fig. 5: Framework of the DGD algorithm

Fig. 6: Group division mechanism

5.2 Group Generating Protocol

If a vehicle has been accessing the VSNs, then the RA will validate the identity of the vehicle. After finishing system initialization, the MFemtocell can help the vehicle automatically interact at the same social/individual hot spot. According to the area size of the hot spot, different cells are generated as shown in Fig. 6. To preserve privacy, we propose the Group Generating Protocol to expand the area of the group constantly, and then a group region is dynamically generated.

Here, we explain the main procedure of group generating. First, when a vehicle v_1 enters a social/individual hot spot, it will monitor whether there are other vehicles in the same hot spot that are sending the Cell-Generate-Request message. If a vehicle v_2 has initiated the request message for cell generating, vehicle v_1 will join the cell. Otherwise, vehicle v_1 initiates the request message. Within a timestamp t_stamp, the cell is finished and the vehicle that first sends the request message becomes the Cell Leader. Second, the Cell Leader broadcasts the Group_Generate_Request message to find the other vehicles. When the number of vehicles in the group is larger than the maximum value K_max (that is, K_t > K_max) or the distance deviation between vehicles is larger than the maximum value d_max (that is, d_t > d_max), then the group is finished. Otherwise, the area of group will be extended to find appropriate vehicles for the time t_stamp. Fig. 6 describes the generative process of the group region. Because the social hot spot is generally an intersection, the trajectory map in the group region is complicated because of different directions of the trace intersection. Therefore, the group division mechanism in the DGD algorithm can more effectively protect the vehicles' trajectory privacy compared with the privacy-preserving scheme MixGroup.
&PK_r, &SK_r, &Cert_r;

7: The vehicle v_i becomes the Cell Leader and waits for other vehicles to join;
8: Cells are finished;
9: Select a cell leader as the group leader by the principle of first come first service;
10: Broadcast the request message Group_Generate_Request;
11: for (K_t <= K_max & & d <= d_max)
12: for (PL (the cell leader nearby group leader) ∈ social hot spot)
13: The cell leader agrees to join group;
14: Calculate K_t and d_i;
15: if (K_t >= K_max || d_i >= d_max)
16: break;
17: end if
18: end for
19: The cell leader (PL ∈ individual hot spot) agrees to join group;
20: Calculate K_t and d_i;
21: end for
22: if (t_Stamp < 0)
23: Group generation fails;
24: end if
25: else
26: Group is finished;
27: Group leader sends the message to the RA;
28: RA judges its legitimacy and distributes GID, PK_G, SK_G, Cert_g for valid groups;
29: end for

As shown in Algorithm 2, group generation consumes a certain amount of time (within the timestamp t_Stamp). Thus, time is the key point for the Group Generating Protocol. If the time setting is shorter, then K_t and d_i are accordingly smaller and cannot reach the requirements of privacy preservation. If the time setting is longer, then procedure of group generation will waste too much time waiting and cannot meet the user’s requirements.

5.3 Pseudonym exchange protocol

To protect the vehicles’ location and trajectory privacy outside of the group area, vehicles periodically send safety messages with pseudonym identities. Vehicles directly communicate with the Base Station. Inside of the group area, vehicles employ group identity (GID). To ensure that the attacker cannot deduce the vehicles’ real identity, this paper increases the pseudonym exchanging opportunities. When a vehicle enters a group region, the vehicle will continuously exchange pseudonyms with other vehicles until the vehicle leaves the group region. A condition for exchanging pseudonyms between vehicles is observed. Assume that each vehicle has the probability of being tracked by a malicious attacker denoted by p_r. Thus, the pseudonym entropy H for the vehicles {v_1, v_2, ⋯, v_k} in a group region can be expressed as follows:

$$H = - \sum_{i=1}^{k} p_i \log_2 p_i$$

When the vehicle’s pseudonym is changed, the probability p_r and the pseudonym entropy H are changed. Therefore, this paper exploits the entropy of the group to exchange the vehicles’ PIDs to preserve privacy.

The pseudo code of the Pseudonym Exchanging Protocol is shown in Algorithm 3. Here, we consider the privacy preservation of vehicle v_i in group region Gr. Before exchanging the pseudonym of vehicle v_j, we first calculate the pseudonym entropy of the Gr, which is denoted by the symbol H_before. Thus, H_before is the entropy before exchanging. In the valid anonymity duration T, if vehicle v_i receives the message pseudonym_exchange_ Request from a vehicle v_j, we will estimate the pseudonym entropy of the group region Gr denoted by the symbol H_after. Of course, the vehicle v_i also belongs to the group region Gr. If the pseudonym entropy H_after is greater than H_before, the two vehicles v_i and v_j will interexchange pseudonyms. In contrast, vehicle v_i will abandon the opportunity of exchange because smaller pseudonym entropy corresponds to lower the level of privacy protection.

In summary, for privacy preservation in the Pseudonym Exchanging Protocol, we calculate the pseudonym entropy H before two vehicles exchange pseudonyms. If H is increased by exchanging pseudonyms, then the two vehicles will successfully complete the process of pseudonym exchange. Otherwise, the exchange will be abandoned.

**Algorithm 3: Pseudonym Exchanging Protocol**

**Input:** Gr, GID, T

**Output:** H_after, PID_r, PK_r, SK_r, Cert_r

1: for (a vehicle v_j ∈ Gr)
2: Calculate the pseudonym entropy H_before; // H_before is the entropy before exchanging.
3: if (T != 0 && vehicle v_j receives the message pseudonym)
4: _exchange_ Request from a vehicle v_j; // (v_j ∈ Gr)
5: Calculate the pseudonym entropy H_after; // H_after is the entropy after exchanging.
6: if (H_after - H_before > 0)
7: Exchange pseudonyms <v_j, v_i>;
8: end if
9: else
10: Abandon the exchange;
11: end if
12: else
13: Send the request message pseudonym_exchange
14: _Request and wait the exchanging;
15: Group Leader sends the exchanged pseudonyms to RA;
16: end for
5.4 Group Revocation

When a vehicle leaves the group region, it will send a message about leaving the group to the Group Leader. If the lifetime of the vehicle is zero \((T_{\text{life}}=0)\), then the vehicle will automatically apply to leave the group. When the number of vehicles in a group is less than the minimum value \(K_{\text{min}}\) (that is, \(K_t < K_{\text{min}}\)) or the distance deviation between vehicles is less than the minimum value \(d_{\text{min}}\) (that is, \(d_i < d_{\text{min}}\)), we determine that the existence of the group is meaningless. Thus, the group will be revoked by the RA. Algorithm 4 presents the pseudo code of the Group Revocation Protocol.

Algorithm 4: Group Revocation Protocol

**Input:** Group region, GID, \(T_{\text{life}}\), \(K_{\text{min}}\), \(d_{\text{min}}\)

**Output:** Resources of group

1. if (a vehicle \(v_i \in G\) Group region \&\& \(T_{\text{life}} < 0\))
2. Initiate the message leaving_group_request;
3. Calculate \(K_t\) and \(d_i\);
4. if \((K_t < K_{\text{min}}) \text{ or } d_i < d_{\text{min}}\)
5. Group leader sends message of group revocation to RA;
6. RA recovers the corresponding resources of the group;
7. end if
8. end if

6. Simulation and Results

In this section, we conduct extensive simulations and evaluate the performance and effectiveness of our proposed DGD algorithm under city and suburban scenarios. We first describe the simulation environment for the city and suburban scenarios. The simulation results from three aspects (anonymity set size \(K\), distance deviation \(D\) and pseudonym entropy \(H\)) are then presented in charts. Finally, through the analysis of the simulation, we show that our proposed DGD algorithm can effectively protect the location and trajectory privacy in 5G-based VSNs.

6.1 Simulation Environment

<p>| Table 1: Simulation parameters |</p>
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vehicle Speed</td>
<td>30 km/h</td>
</tr>
<tr>
<td>Vehicle density</td>
<td>0.3 v/m for city scenario</td>
</tr>
<tr>
<td>Vehicle density</td>
<td>0.2 v/m for suburban scenario</td>
</tr>
<tr>
<td>Anonymity Duration</td>
<td>600 s</td>
</tr>
<tr>
<td>(K_{\text{min}})</td>
<td>35</td>
</tr>
<tr>
<td>(d_{\text{min}})</td>
<td>600 m</td>
</tr>
<tr>
<td>(K_{\text{max}})</td>
<td>50</td>
</tr>
<tr>
<td>(d_{\max})</td>
<td>750 m</td>
</tr>
</tbody>
</table>

We first simulate a graph to represent a large real geographic area \((3000 \text{ m} \times 3000 \text{ m})\) using the Generic Mobility Simulation Framework, GMSF\(^{30}\). For emulating the performance of MFemtocell, we add certain functions to the GMSF simulator that are compliant with 3GPP LTE specifications\(^{31}\). Then, we divide the graph into 10×10 small cells. Each cell is labeled a social hot spot or individual hot spot according to the real geographic area. To determine whether to exchange pseudonyms, each vehicle is labeled with an ordered pair \(<\text{PID}_i, p_i>\), where \(\text{PID}_i\) represents the vehicle identity and \(p_i\) is the tracked probability.

As we know, GMSF supports the import of vehicle traces for three different areas (city, suburban and rural). For simulating the different road information in Chengdu City, Sichuan Province, China, we design a city scenario and a suburban scenario. The only difference between the two scenarios is the vehicle density. We set the vehicle density of the city scenario as 0.3 v/m (that is, 900 vehicles in 3000 m × 3000 m area), whereas we set the vehicle density of the suburban scenario as 0.2 v/m (that is, 600 vehicles in 3000 m × 3000 m area). Table 1 shows the simulation parameters in details.

6.2 Simulation Results

A. Location Privacy

To simplify our experiment, we assume that a vehicle is in the social hot spot and sends a request message for group generating. We analyze the location privacy in the city scenario and suburban scenario from two aspects: anonymity set size and distance deviation.

Fig. 7: Anonymity set size of a group region for the city scenario

Fig. 8: Anonymity set size of a group region for the suburban scenario
In Fig. 7, we measure the anonymity set size \((K_t)\) of a single group region within 100s in the city scenario, which shows that the anonymity set size with the DGD algorithm is larger than the MixGroup algorithm when the time \(t \leq 80\) s. Our proposed DGD algorithm approaches the maximum \((K_{\text{max}}=50)\) at \(t=70\) s, whereas the MixGroup algorithm achieves the \(K_{\text{max}}\) value at \(t=90\) s. Evidently, with the assistance of MFemtocell, the DGD algorithm possesses a faster convergence rate. Hence, from perspective of protecting the location privacy in a city scenario, the DGD algorithm can perform better compared with the MixGroup algorithm.

However, the MixGroup algorithm approaches the maximum \((K_{\text{max}}=50)\) in a lower density area in the suburban scenario. As shown in Fig. 8, although the DGD and MixGroup algorithm obtain their peak almost simultaneously (at approximately \(t=80\) s), the anonymity set size in the DGD algorithm is much larger than that of the MixGroup algorithm. A larger anonymity set size corresponds to a greater intensity of privacy preserving. Thus, our proposed DGD algorithm shows its superior performance of protecting location privacy in the suburban scenario.

Fig. 9 compares the anonymity set size between the city scenario and suburban scenario in the DGD algorithm. Because the vehicle density is higher in the city scenario than the suburban scenario, the DGD algorithm obtains the maximum anonymity \((K_{\text{max}}=50)\) at \(t=70\) in the city scenario and at \(t=90\) in suburban scenario. Therefore, the DGD algorithm can protect location privacy well in both the city scenario and the suburban scenario.

Fig. 10 compares the distance deviation of a single group region between the DGD and MixGroup algorithm in the city scenario. Fig. 10 shows that the convergence rate of the DGD algorithm is faster than that of the MixGroup algorithm. The DGD algorithm achieves the convergence value at approximately \(t=60\) s, whereas the MixGroup algorithm achieves it at \(80\) s. This finding further demonstrates that our proposed DGD algorithm has faster response speeds and effectively protects location privacy under the city scenario.

The distance deviation of a group region for the suburban scenario is shown in Fig. 11. Two rapid increases are observed at \(t=30\) s and \(t=60\) s separately in MixGroup algorithm, while the growth of distance deviation shows a smooth increase in our proposed DGD algorithm. These data reveal that there are two social hot spots in the MixGroup algorithm, which may be caused by the leaking of location privacy, whereas the DGD algorithm performs better in preserving location privacy.

Fig. 12 compares the distance deviation of a group region between the city and suburban scenarios. We find that the distance deviation is smaller in the city scenario than in the suburban scenario when \(t<60\) s. However, the distance deviation of the city scenario overlaps the suburban scenario on the end. Both scenarios realize location privacy preservation when the distance deviation falls between 750 m \((d_{\text{max}})\) and 600 m \((d_{\text{min}})\).
B. Trajectory Privacy

To better measure the trajectory privacy, we also assume that a vehicle starts in a social hot spot and sends a safety message periodically with a speed of 30 km/h. Then, a trajectory is generated that includes 6 group regions in the anonymity duration (T=600 s). Here, the trajectory privacy is described with the average anonymity set size, average distance deviation and pseudonym entropy.

Fig. 13 shows the anonymity set size of the trajectory under the city scenario. Comparing the DGD and MixGroup algorithm, the $K$ in the DGD algorithm is larger than that of the MixGroup overall. Through formula 1, we can calculate the average anonymity set size as $K=42.33$ in the MixGroup algorithm, whereas the $K=47.20$ in the DGD algorithm. Hence, our proposed algorithm can effectively protect the trajectory privacy in the city scenario.

In the MixGroup algorithm for the suburban scenario, three types of time settings {200 s, 500 s, and 600 s} are observed in which the anonymity set size is below the minimum ($K_{\text{min}}=35$). Fig. 14 shows that the average anonymity set size $K$ is 34.63. However, the anonymity set size of the DGD algorithm is close to the $K_{\text{max}}$, and the average anonymity set size $K$ ($K=43.62$) in the DGD algorithm is much larger than that in the MixGroup algorithm. Thus, we find that the DGD algorithm effectively provides the service of trajectory privacy preservation in the suburban scenario.

Fig. 15 compares the anonymity set size on the trajectory between the city scenario and suburban scenario. In general, the anonymity set size in the city scenario is larger than that in the suburban scenario. Based on the above, the performance of our proposed DGD algorithm for trajectory privacy preservation in the city scenario is superior to that of the suburban scenario.

Fig. 16-18 show the distance deviation of the trajectory in the anonymity duration of 600 s. Based on formula 5, the average distance deviation of the DGD algorithm is $D=727.67$ while that of the MixGroup algorithm is 621.33 in the city scenario (as shown in Fig. 16). In suburban scenario (see Fig. 17), we find that the average distance deviation $D$ is 609.17 using the MixGroup algorithm and $D=723.00$ with the DGD algorithm. Thus, the DGD algorithm, which presents a higher average distance deviation, can effectively protect the trajectory privacy both in the city and suburban scenarios.
We compare the distance deviation between the city scenario and the suburban scenario in Fig. 18. Generally, the distance deviation in the city scenario is larger than that in the suburban scenario. Accordingly, the DGD algorithm is more effective in the city scenario in terms of preserving the trajectory privacy.

Fig. 19 compares the pseudonym entropy on the trajectory between the DGD and MixGroup algorithms in the city scenario, and it shows that the pseudonym entropy with the DGD algorithm is larger than that with the MixGroup algorithm. The degree of privacy protection is directly proportional to the pseudonym entropy. From Fig. 20, the same result is found in the suburban scenario. Thus, in both the city scenario and suburban scenario, the DGD algorithm has a strong ability to protect trajectory privacy in the 5G-based VSNs compared with the MixGroup algorithm.

7. CONCLUSIONS

In this paper, we studied the problem of protecting location and trajectory privacy in 5G-based VSNs. To dynamically divide the group region and meet the high real-time demands of users, we propose a system model of 5G-based VSNs that applies MFemtocell technology. In our proposed system model, we design an efficient DGD algorithm to protect a vehicle’s location and trajectory privacy. The DGD algorithm comprises four stages: system initialization, group generating, pseudonym exchange, and group cancellation. Through the simulations, we show that, compared with existing solutions for generating the group region, our algorithm reduces the time delay and effectively protects the users’ location and trajectory privacy in 5G-based VSNs.

ACKNOWLEDGEMENT

This research was partially supported by the National Natural Science Foundation of China (61571098), Fundamental Research Funds for the Central Universities (ZYGX2016J217).

REFERENCES

Global Communications Conference (GLOBECOM), 1-7, 2016.


