A Trajectory Substitution Privacy Protection Scheme in location-based services

Cheng Song, Yadong Zhang*, Xinan Gu, Lei Wang and **Zhizhong Liu**
School of Computer Science and Technology, Henan Polytechnic University
Jiaozuo, Henan, 454000, China
[e-mail: songcheng@hpu.edu.cn]
[e-mail: 18339161026@163.com]
*Corresponding author: Yadong Zhang

Received November 26, 2018; revised March 17, 2019; accepted April 9, 2019; published September 30, 2019

Abstract

Aimed at the disclosure risk of mobile terminal user's location privacy in location-based services, a location-privacy protection scheme based on similar trajectory substitution is proposed. On the basis of the anonymized identities of users and candidates who request LBS, this scheme adopts trajectory similarity function to select the candidate whose trajectory is the most similar to user's at certain time intervals, then the selected candidate substitutes user to send LBS request, so as to protect user's privacy like identity, query and trajectory. Security analyses prove that this scheme is able to guarantee such security features as anonymity, non-forgeability, resistance to continuous query tracing attack and wiretapping attack. And the results of simulation experiment demonstrate that this scheme remarkably improve the optimal candidate' trajectory similarity and selection efficiency.

Keywords: similar trajectory, privacy protection, query substitution, LBS

This work was supported by the National Natural Science Foundation of China (61872126, 61772159, 61300216); the Science and Technology Research Program of He Nan Province (192102210123, 182102110333, 172102310677).

http://doi.org/10.3837/tiis.2019.09.024 **ISSN : 1976-7277**

1. Introduction

Along with the development and improvement of intelligent mobile devices, GPS and wireless communications technology, location-based services (LBS) [1-3] has been gaining users' attention. For instance, mobile terminal users may send location services request to LBS server to acquire the information on weather, food & beverage, navigation, etc., so as to enjoy the conveniences of LBS. However, in this process, users' private information is liable to be leaked, such as users' living habit, interests or hospitalization information. Therefore, while enjoying the conveniences brought by LBS, how to protect users' privacy proves to be one of the urgent problems to be solved in this field[4-6]. Up to now, despite some scholars have made certain progress in protecting the disclosure of users' instant location information[7], attackers are still able to obtain users' relevant privacy information by analyzing users' location trajectories. As a result, on the basis of protecting users' location privacy, how to further protect the disclosure of users' trajectory information turns out to be a hot issue in current research[8].

Aiming to address the deficiencies in the existing schemes, this paper proposes a location privacy preservation scheme based on similar trajectory's substitute query. In this scheme, trusted anonymous server (TAS) is employed to anonymize the identities of all users requesting LBS, so when a certain user sends request for location service, TAS adopts trajectory similarity algorithm to calculate the trajectory similarity value of all candidates and initiators within user's specific area at certain time intervals, from which an optimal candidate with trajectory most similar to the requested user's trajectory is selected to substiture real user in requesting LBS. Consequently, user's privacy like identity, query and trajectory can be guaranteed. This paper also employs random oracle model (ROM) to verify the scheme's security like anonymity and non-forgeability, and to analyze such security properties as resistance to continuous query tracing attack and resistance to wiretapping attack. We also conduct simulation experiment on the optimal candidate's trajectory similarity and selection algorithm, which demonstrates the effectiveness and efficiency of this scheme. The following summarizes the main contributions of this paper:

1. We propose a location privacy preservation scheme based on similar trajectory's query substitution. In this scheme, the real user is replaced by the optimal candidate to submit the LBS service request. As a result, the trajectory of real user is protected.

2. Through anonymizing the identities of user and candidates who request LBS, calculating the trajectory similarity between all candidates and the initiator within a certain period of time by using trajectory similarity algorithm, and choosing the candidate with the optimal trajectory similarity to replace real user in requesting LBS, the real user's private information about identity and location is effectively protected.

3. We conduct security analyses of the proposed solution. Security analyses prove that this scheme is able to guarantee such security features as anonymity, non-forgeability, resistance to continuous query tracing attack and wiretapping attack. And the results of simulation experiment demonstrate that this scheme remarkably improve the optimal candidate' trajectory similarity and selection efficiency.

The rest of this paper is organized as follows: In section 1, we show the Introduction. In section 2, we introduce the related work. In section 3, we introduce the preliminaries. The trajectory privacy protection scheme is described in detail in section 4. We offer the security

analyses in section 5. In section 6, we give the simulation experiment about the similarity of the optimal candidate's trajectory. The conclusions are given in section 7.

2. Related Work

In recent years, in view of protecting users' trajectory privacy, some scholars home and abroad have done lots of research and achieved certain positive results. Pan X et al.[9] proposed a distortion-based anonymity scheme for mobile users' continuous queries, in which a generalized region where at least K number of users including the user who initiates the request is adopted and the generalized region is measured in reference to users' motion velocity and direction in order to preserve users' trajectory privacy. Later, Freudiger J et al.[10] proposed a metric based on the mobility profiles of mobile nodes to evaluate the mixing effectiveness of possible mix zone locations. As the location privacy achieved with mix zones depends on their placement in the network, this scheme analyzes the optimal placement of mix zones with combinatorial optimization techniques, thus succeeding in protecting users' trajectory privacy. Kato et al. [11] assumed that users' movements are known in advance, and on the basis of users' mobility trajectories and pauses, selected each hop of dummy trajectories according to the deviation angles of users' locations and accessibility. Finally they proposed a dummy-based anonymization method based on user trajectory with pauses. However, in this scheme the similarity of dummy trajectories selected still needs to be improved. In reference[12], Huo Z et al. proposed a trajectory privacy-preserving method in mobile social network services. In their method, firstly they buffer the check-in sequences of pseudonym users, and then build prefix trees for buffered check-in sequences, prune and re-construct prefix trees to get the k-anonymized version. Finally they traverse the k-anonymized prefix tree to get k-anonymized check-in sequences, which can achieve a privacy guarantee of k-anonymity. In reference[13], users employ a buffer to record their history locations and the exchanged pseudo-locations. When requesting LBS, users choose from the buffer and submit k number of locations (including their own locations), so as to achieve the trajectory privacy security of $K-$anonymized users.

Chen R et al. [14] proposed a tailored privacy model to anonymize trajectory data by means of local suppression, which serves to preserve users' trajectory privacy. Afterwards, Zhao J et al. [15] put forward a novel privacy preserving method applying to trajectory data publishing, which functions to prevent identity linking attack as well as attribute attack. Li F H et al. [16] proposed an efficient scheme for trajectory privacy protection, which constructs $(k-1)$ dummy trajectories based on side information, trajectory similarity and users' mobility pattern, thus attackers fail to identify the real trajectory, then the terminal trajectory privacy can be guaranteed. Ye A et al. [17] proposed a location privacy-preserving scheme based on l-queries, in which some fake queries will be randomly injected into real continuous queries, so as to reduce temporal relevancy of continuous queries, then protect the privacy of users' identities, locations and query content. In the location privacy protection approach put forward by Sun G et al. [18], location labels are adopted to distinguish mobile users' sensitive locations and ordinary locations, both of which are protected at different levels, thus the response efficiency of LBS requests is improved. According to Li et al. [19], to improve the efficiency of selecting dummy locations, an credit-incentive mechanism is introduced in K-anonymity scheme. Based on fuzzy logic, each user's credit level corresponds to certain probability threshold value. A user can get the help of other users only on the condition that his credit level matches certain probability threshold value. In this way, users are motivated to actively help others in constructing K-anonymity. In a sense, all the schemes mentioned above invariably

originate from the location privacy protection scheme for single-time LBS, and thus ignore two issues: (1) security of communications messages in users' LBS request; (2) certain location relevancy in continuous query may lead to the disclosure of users' real location information.

3. Preliminaries

3.1 System model of trajectory privacy preservation

As is shown in **Fig. 1**, the system model of trajectory privacy preservation is mainly composed of three entities: mobile terminal (MT), trusted anonymous server (TAS) and LBS server. Each entity functions as follows:

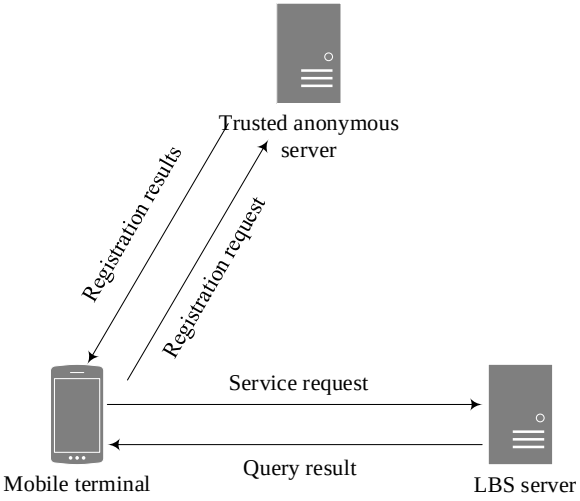

Fig. 1. System architecture of trajectory privacy preservation

(1) MT: MT has two functions: sending request for anonymization to TAS, then verifying the validity of the anonymity and generating corresponding key information; sending location query request to LBS server and receiving query result.

(2) LBS server: as the core of location privacy preservation system, it processes location service request for anonymity from MT and returns query results to MT.

(3) TAS: it records attribute matrix information of the registered user (time sampling and the corresponding location coordinate sampling), processes request for anonymity from MT, and selects user with the optimal trajectory similarity for MT, then generates and releases system-related parameters.

3.2 Bilinear pairings

Definition [20] : suppose $(G_1,+)$, $(G_2,*)$ are separately the addition cycle group and multiplication cycle group of prime order q, then bilinear pairing $e:G_1\times G_1\to G_2$ satisfies the following conditions:

(1) Bilinearity: Bilinearity means that the function $G_1(P,Q)$ is linear to both the variables P and Q. For example, for $\forall a,b\in Z_q^*, P,Q\in G_1$, $e(aP,bQ)=e(P,Q)^{ab}=e(abP,Q)$.

(2) Non-degeneracy: $\exists P,Q \in G_1$, satisfies $e(P,Q) \neq 1_{G_2}$, 1_{G_2} signifies the identity element of group $(G_2,*)$.

(3) Symmetry: $\forall P,Q \in G_1$, $e(P,Q)=e(Q,P)$.

(4) Computability: $\forall P,Q \in G_1$, there being polynomial time algorithm which can calculate $e(P,Q)$.

3.3 Trajectory Similarity

Trajectory similarity denotes the similarity degree of two mobility trajectories, and is usually measured with trajectory similarity function. Let user's real mobility trajectory be $T_u = \{ID_A,(x_0,y_0,t_0),(x_1,y_1,t_1),\cdots,(x_n,y_n,t_n)\}$, in which, ID_A signifies the identity of anonymized user, (x_i,y_i,t_i) means the location of t_i at the moment ID_A is (x_i,y_i), and sampling time t_i satisfies $t_0 < t_1 <,\cdots,< t_n$. Suppose user's direction of mobility trajectory from initial point t_0 to moment t_i is ϕ_i, then $\tan\phi_i = ((y_i - y_0)/(x_i - x_0))$, that is, $\phi_i = \arctan((y_i - y_0)/(x_i - x_0))$. Consequently, mobility trajectory T_u can be shown as: $T_u = \{(x_0,y_o,t_0),\langle(\phi_1,t_1),(\phi_2,t_2),\cdots,(\phi_n,t_n)\rangle\}$. Similarly, the mobility trajectory of potential candidate is $T_C = \{ID_C,(x_0^c,y_0^c,t_0^c),(x_1^c,y_1^c,t_1^c),\cdots(x_n^c,y_n^c,t_n^c)\}$, which can be shown as $T_C = \{(x_0^c,y_0^c,t_0^c),\langle(\phi_1^c,t_1^c),\cdots,(\phi_n^c,t_n^c)\rangle\}$, while

$$\phi_i^c = \arctan\frac{y_i^c - y_0^c}{x_i^c - x_0^c}, (1 \le i \le n, 1 \le c \le k-1). \quad (1)$$

The similarity degree of mobility trajectories is:

$$\sigma^2 = ((\sum_{i=1}^{n}\frac{\phi_i^c - \phi_i}{2\pi})/n)^2 \quad (2)$$

Obviously, $\sigma^2 \in [0,1]$, and the less σ^2 is, the stronger the similarity between user's mobility trajectory and candidate's trajectory.

4. Location Privacy Preservation Scheme

This scheme mainly consists of four phases: system initialization, registration, calculation of trajectory similarity, and location service request of query substitution.

4.1 System Initialization

In this phase, system parameters are generated as follows:

Step 1: Select two cyclic groups $(G_1,+)$ and $(G_2,*)$ with prime q as the order number, and bilinear map $e: G_1 \times G_1 \to G_2$, and P is a generator of G_1.

Step 2: Define three secure harsh functions: $H_1:\{0,1\}^* \to G_1^*$, $H_2:\{0,1\}^* \to \{0,1\}^n$ and $H_3: G_2 \to \{0,1\}^n$, in which n denotes a positive integer, while $\{0,1\}^*$ the binary strings at any length.

Step 3：TAS selects two random number $s \in Z_q^*$ and $Q \in G_1^*$, calculates their public key $PK_{anon} = sP$ and secret key $SK_{anon} = sQ$, in which Z_q^* signifies the integers' multiplication group of module q.

Step 4：TAS publicizes the system parameter: $\{G_1, G_2, e, k, n, P, PK_{anon}, Q, H_1, H_2, H_3\}$.

4.2 Registration

In this phase, TAS helps to anonymize user's identity and generates corresponding parameters, as follows:

Step 1: User U sends its real identity ID safely to TAS via MT.

Step 2: TAS randomly generates a $m \times n$ matrix χ ($2 \le m < n$) and a m-dimensional column vector $\boldsymbol{\rho}$, while the rank of coefficient matrix χ is equal to that of augmented matrix $\overline{\chi}$, that is, $R(\chi) = R(\overline{\chi})$. Obviously, $R(\chi) < n$, then linear equation $\boldsymbol{\chi d = \rho}$ has infinite solutions.

Step 3：TAS distributes a unique n-dimensional column vector $\boldsymbol{d}_i$ for each registered user, and $\boldsymbol{d}_i$ satisfies $\boldsymbol{\chi d_i = \rho}$, that is, $\boldsymbol{d}_i$ is a solution of linear equation $\boldsymbol{\chi d_i = \rho}$. Then, TAS randomly selects a n-dimensional column vector $\boldsymbol{D}$, calculates the fake identities $PID_U = \boldsymbol{D}^T \bullet \boldsymbol{d}_i$ of user U, as well as the relevant verified parameters $Q_u = H_1(PID_u)$ and $S_U = sQ_U$, finally TAS returns $\{PID_U, S_U, d_i\}$ to user U via secure channel.

Step 4：After receiving message $\{PID_U, S_U, d_i\}$, user U calculates $\tilde{Q}_u = H_1(PID_u)$ and judges whether $e(S_U, P) \stackrel{?}{=} e(\tilde{Q}_U, PK_{anon})$ is valid or not. If valid, user U randomly selects a secret value $r_U \in Z_q^*$, calculates its public key $PK_U = r_U PK_{anon}$ and secret key $SK_U = r_U S_U$; if not valid, return to **Step 1**.

4.3 Calculation of Trajectory Similarity

In this phase, TAS employs trajectory similarity function to calculate the similarity between all candidates' trajectories and the trajectory of user U (to calculate successively the trajectory similarity of the initiator and the candidate), as follows:

Step 1：Assume the sampling point of mobile terminal user U's real trajectory and candidate's mobility trajectory respectively is $T_U = \{ID_U, (x_0, y_0, t_0), (x_1, y_1, t_1), \cdots, (x_k, y_k, t_k)\}$ and $T_C^i = \{ID_C, (x_0^c, y_0^c, t_0^c), (x_1^c, y_1^c, t_1^c), \cdots (x_k^c, y_k^c, t_k^c)\}$.

Step 2： Based on $\phi_i = \arctan((y_i - y_0)/(x_i - x_0))$ （$1 \le i \le n$）, TAS successively calculates the motion angle ϕ_i of candidate's and user U's trajectories from initial moment t_0 to moment t_i, then the mobility trajectory of user U (T_U) and that of candidate (T_C) can be respectively shown as:

$T_U = \left\{(x_0, y_o, t_0), \left\langle(\phi_1, t_1), (\phi_2, t_2), \cdots, (\phi_n, t_n)\right\rangle\right\}$ and $T_C = \left\{(x_0^c, y_0^c, t_0^c), \left\langle(\phi_1^c, t_1^c), \cdots, (\phi_n^c, t_n^c)\right\rangle\right\}$.

Step 3：Based on trajectory similarity formula: $\sigma^2 = (\frac{\sum_{i=1}^{n} \frac{\phi_i^c - \phi_i}{2\pi}}{n})^2$, TAS calculates the similarity degree σ^2 between candidate's trajectory T_C and mobile user U's trajectory T_U.

Step 4: Repeat **Step1** to **Step 3** to calculate the similarity between other candidates' trajectory and user U 's trajectory. Then select the candidate with minimum trajectory similarity as the optimal candidate. If there is more than one candidate with minimum similarity, then select the candidate which is farther in distance from user U as the optimal candidate according to Euclidean distance formula:

$$dis(T_U, T_i) = \frac{\sum_{i=1}^{n} \sqrt{(x_U - x_i)^2 - (y_U - y_i)^2}}{n}. \quad (3)$$

4.4 Location Service Request of Query Substitution

In this phase, the selected candidate with optimal trajectory similarity will substitute user U to request LBS service，:

Step 1：User U launches broadcast to obtain candidate's anonymous identity $m_1 = \{PID_C^1, PID_C^2, \cdots, PID_C^{k-1}\}$ within the range of communications, randomly selects $\omega \in Z_q^*$, calculates $Q_{LBS} = H_1(ID_{LBS})$, $c_1 = \langle \omega P, m_1 \oplus H_3(e(Q, PK_{anon})^{\omega}) \rangle = \langle E, F \rangle$ and $c_2 = \langle \omega P, m_2 \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega}) \rangle$, in which ID_{LBS} signifies the identity label of LBS server, $m_2 = \{L_U, M_U, K_s\}$, L_U user's location, M_U the query content, K_s the session key of user and LBS server, and $PK_{LBS} = r_{LBS} sP$ the public key of LBS server; then sends data packet $Meg_{UoN} = \{c_1, c_2\}$ to TAS.

Step 2：Receiving the request, TAS randomly selects $i \in Z_q^*$, calculates $I = iZ$ and $r = ip$, then sends the message $\{t_1, I, H_2(r \| ID_{anon} \| t_1)\}$ to user U , in which t_1 is the time stamp, and ID_{anon} is the identification of TAS.

Step 3：After receiving the message from TAS, user U firstly calculates $R = Id_i$, then verifies the equation $H_2(R \| ID_{anon} \| t_1) \stackrel{?}{=} H_2(r \| ID_{anon} \| t_1)$. If valid, sends $\{t_2, H_2(R \| ID_{anon} \| t_1 \| t_2)\}$ to TAS, in which t_2 is the time stamp.

Step 4：Receiving the message, TAS verifies the equation $H_2(R \| ID_{anon} \| t_1 \| t_2) \stackrel{?}{=} H_2(r \| ID_{anon} \| t_1 \| t_2)$. If valid, calculates $F \oplus H_3(e(SK_{anon}, E))$ to get message m_1 ; extracts the information of all candidates' attribute matrix according to message m_1 , employs trajectory similarity algorithm to select the optimal candidate B within a certain time interval, randomly selects $\beta \in Z_q^*$, and calculates $Q_B = H_1(PID_B)$ and $c_3 = \langle \beta P, m_3 \oplus H_3(e(Q_B, PK_B)^{\beta}) \rangle$, in which $m_3 = \{PID_U\}$, $PK_B = r_B sP$ is the public key of user B ; finally sends data packet $Meg_{NoB} = \{c_2, c_3\}$ to user B , and B will substitute user U to initiate LBS request.

Step 5: Receiving the message, user B calculates $m_3 \oplus H_3(e(Q_B, PK_B)^{\beta}) \oplus H_3(e(SK_B, \beta P))$ to get message m_3 , in which $SK_B = r_B sQ_B$ is the private key of user B , then sends data packet $Meg_{UoN} = \{c_2\}$ to LBS server.

Step6: Receiving the message, LBS server calculates $m_2 \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega} \oplus H_3$ $(e(SK_{\text{LBS}}, \omega P))$ to get message m_2 , then adopts m_2 to obtain query result $m_4 = \{MEG\}$, finally

sends data packet $Meg_{SoB} = \{En_{K_s}(m_4)\}$ to user B, in which $En_K()$ is the encrypted symmetric function, and MEG is the query result of LBS service request.

Step 7: Receiving message $Meg_{SoB} = \{En_{K_s}(m_4)\}$, user B verifies the identity of user U. If valid, then sends data packet $Meg_{BoU} = \{En_{K_s}(m_4)\}$ to user U; otherwise, cease the service.

Step 8: After receiving message $Meg_{BoU} = \{En_{K_s}(m_4)\}$, user U decrypts it to obtain the query result MEG.

5. Scheme Analyses

This paper analyzes the scheme in terms of its correctness and security.

5.1 Correctness Analysis

This process demonstrates the correctness of message encryption and decryption between user and TAS, user and user, as well as user and LBS server, that is, the correctness of encryption and decryption of messages m_1, m_2 and m_3.

(1) It is known that user U adopts the public key of TAS to transform the anonymous identity m_1 of candidates within the range of communications into ciphertext $c_1 = \langle \omega P, m_1 \oplus H_3(e(Q, PK_{anon})^{\omega}) \rangle = \langle E, F \rangle$, so after receiving the ciphertext, TAS adopts secret key SK_{anon} to obtain plain text m_1 by calculating $F \oplus H_3(e(SK_{anon}, E))$, which is demonstrated as follows:

$$\begin{aligned} F \oplus H_3(e(SK_{anon}, E)) &= F \oplus H3(e(sQ, \omega P)) \\ &= F \oplus H_3(e(Q, sP)^{\omega}) \\ &= F \oplus H_3(e(Q, PK_{anon})^{\omega}) \\ &= m_1 \end{aligned}$$

(2) It is known that user U adopts LBS server's public key to transform the requested data message m_2 into ciphertext $c_2 = \langle \omega P, m_2 \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega}) \rangle = \langle \varphi, \eta \rangle$, then after receiving the ciphertext, LBS adopts secret key $SK_{LBS} = r_{LBS} sQ_{LBS}$ to obtain plain text m_2 by calculating $\eta \oplus H_3(e(SK_{\mathrm{LBS}}, \varphi))$, which is demonstrated as follows:

$$\begin{aligned} \eta \oplus H_3(e(SK_{\mathrm{LBS}}, \varphi)) &= \eta \oplus H_3(e(r_B S_{\mathrm{LBS}}, \omega P)) \\ &= \eta \oplus H_3(e(r_{LBS} sQ_{LBS}, \omega P)) \\ &= \eta \oplus H_3(e(Q_{LBS}, r_{LBS} sP)^{\omega}) \\ &= \eta \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega}) \\ &= m_2 \end{aligned}$$

(3) It is known that user U employs the public key PK_B of the optimal candidate B to transform its fake identity m_3 into ciphertext $c_3 = \langle \beta P, m_3 \oplus H_3(e(Q_B, PK_B)^{\beta}) \rangle = \langle Z = \beta P, T \rangle$, then after receiving the ciphertext, the optimal candidate B adopts the secret key $SK_B = r_B S_B = r_B sQ_B$ to obtain the plain text m_3 by calculating $T \oplus H_3(e(SK_B, Z))$, which is demonstrated as follows:

$$T \oplus H_3(e(SK_B, Z)) = T \oplus H_3(e(r_B S_B, \beta P))$$

$$
\begin{aligned}
&= T \oplus H_3(e(r_B sQ_B, \beta P)) \\
&= T \oplus H_3(e(Q_B, r_B sP)^{\beta}) \\
&= T \oplus H_3(e(Q_B, r_B S_B)^{\beta}) \\
&= T \oplus H_3(e(Q_B, PK_B)^{\beta}) \\
&= m_3
\end{aligned}
$$

To sum up, the encryption and decryption process of the proposed scheme is correct.

5.2 Communication Cost

First of all, we consider the communication overhead between the user and TAS. The data packet sent by the user to TAS is $Meg_{UoN} = \{c_1, c_2\}$. The size of the encrypted data c_1 varies according to the number of candidates k, and the size of the encrypted data packet c_2 is constant. So its communication overhead is $O(k)$. Then we consider the communication overhead between the user and TAS. The message that the user queries to TAS is $Meg_{NoB} = \{c_2, c_3\}$, whose size is constant. In addition, we also consider the communication overhead between the optimal candidate and LBS server. The message that the candidate queries to LBS server is $Meg_{UoN} = \{c_2\}$, whose size is constant. The query results presented to the candidate is $Meg_{SoB} = \{En_{K_s}(m_4)\}$. Its size varies with the number M of Points of Interest (POIs). So the communication overhead between the optimal candidate and LBS server is $O(M)$. Finally, we consider the communication overhead when the optimal candidate presents the query to the user. The query result is $Meg_{BoU} = \{En_{K_s}(m_4)\}$. Its size varies with the number M of Points of interest (POIs). So the communication overhead between the user and the optimal candidate is $O(M)$.

In order to analyze the computational overhead of the scheme qualitatively, we use T_{par} to denote a bilinear pairing operation, T_h to denote a hash function operation and T_{mul} to denote a point multiplication operation. Other operations are ignored here for being relatively simple and lower time consumption. We first consider the user's computational overhead. According to the formulas $Meg_{UoN} = \{c_1, c_2\}$ 、 $Q_{LBS} = H_1(ID_{LBS})$ 、 $c_1 = \langle \omega P, m_1 \oplus H_3(e(Q, PK_{anon})^{\omega}) \rangle$ and $c_2 = \langle \omega P, m_2 \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega}) \rangle$, the computational overhead is $2T_{par} + 3T_h + 3T_{mul}$. Then, considering the computational overhead of TAS, according to formulas $Meg_{NoB} = \{c_2, c_3\}$ 、 $Q_B = H_1(PID_B)$ and $c_3 = \langle \beta P, m_3 \oplus H_3(e(Q_B, PK_B)^{\beta}) \rangle$, the computational overhead is $T_{par} + 2T_h + 2T_{mul}$. Then we consider the computational overhead of the optimal candidate. According to the formula $m_3 \oplus H_3(e(Q_B, PK_B)^{\beta}) \oplus H_3(e(SK_B, \beta P))$, the computational overhead is $2T_{par} + 2T_h + 2T_{mul}$. Finally, we calculate the computational overhead of LBS server. According to the formula $m_2 \oplus H_3(e(Q_{LBS}, PK_{LBS})^{\omega} \oplus H_3(e(SK_{\mathrm{LBS}}, \omega P))$, the computational overhead is $2T_{par} + 2T_h + 2T_{mul}$. As is stated, the computational overhead of the entire process is $7T_{par} + 9T_h + 9T_{mul}$.

5.3 Security Analysis

The security analysis of the scheme is conducted in terms of three aspects: anonymity, non-forgeability and resistance to query tracing attack.

5.3.1 Anonymity game

Step 1：The attacker launches query to obtain the system parameters: $\{G_1, G_2, e, k, n, P, PK_{anon}, Q, H_1, H_2, H_3\}$, and the necessary information of the parameters;

Step 2：The attacker selects two totally different messages m_b and m_{1-b} as the ID messages of user U_1 and U_2;

Step 3：Select random bit $b \in \{0,1\}$, then send m_b and m_{1-b} to two users U_1 and U_2, while b is not open to the attacker;

Step 4：For U_1 and U_2, TAS respectively finds the optimal candidates B_1 and B_2, then sends the encrypted ID c_b and c_{1-b} to B_1 and B_2;

Step 5：If the encrypted information c_b and c_{1-b} received by B_1 and B_2 corresponds respectively with message m_b and m_{1-b}, then sends c_b and c_{1-b} to the attacker in random order; otherwise, return $\perp$ to the attacker;

Step 6：If the attacker A decrypts c_b, and is able to output the message $m_b^{'} = m_b$, then he wins the game.

The advantage of attacker winning this game is defined as: $Adv(A) = |Pr[A]|$, in which $Pr[A]$ represents the probability of attacker A outputting the message $m_b^{'} = m_b$, that is, the probability of attacker obtaining user's ID information.

Theorem 1：In the trajectory privacy preservation scheme, assume attacker A is able to win the anonymity game with negligible probability, then this scheme satisfies the requirement for anonymity.

Proof：Assume attacker A as the attacker in the anonymity game in Theorem 1, if $\perp$ is received in Step 5, then attacker A fails to obtain any useful information. Then consider another possibility: assume attacker A obtains two encrypted data packets $M_{NoB_1} = \{c_b\}$ and $M_{NoB_2} = \{c_{1-b}\}$, in which $c_b = \left\langle \upsilon P, m_b \oplus H_3(e(Q_{B_1}, PK_{B_1})^{\upsilon}) \right\rangle$, $c_{1-b} = \left\langle \zeta P, m_{1-b} \oplus H_3(e(Q_{B_{1-b}}, PK_{B_{1-b}})^{\zeta}) \right\rangle$. $\upsilon \in Z_q^*$ and $\zeta \in Z_q^*$ are the random number generated by TAS, while PK_{B_1} and PK_{B_2} are respectively the public keys of user B_1 and B_2.

Let the probability of the attacker obtaining the encrypted information c_b is $P(\mathbb{Q})$, because c_b and c_{1-b} are sent to attacker A in random order, so $P(\mathbb{Q}) = \frac{1}{2}$. Assuming that the probability of the attacker decrypting the message c_b is $\mathrm{P}(\ell)$, the probability that the attacker finally obtains the identity information m_b is $\mathrm{P}(m) = \mathrm{P}(\ell)\,\mathrm{P}(\mathbb{Q}) = \frac{1}{2}\mathrm{P}(\ell)$. Assuming that the attacker's secret key is $SK_{Attack} = r_{Attack} s_{Attack} Q_{Attack}$, if he tries to obtain user's ID information m_b by decrypting the ciphertext $c_b = \left\langle \upsilon P, m_b \oplus H_3(e(Q_{B_1}, PK_{B_1})^{\upsilon}) \right\rangle$, then he must find the solution of $H_3(e(Q_{B_1}, PK_{B_1})^{\upsilon})$. Attacker can assume $Q_{Attack} = Q_{B_1}$ and $r_{Attack} s_{Attack} P = PK_{B_1}$, and calculate $H_3(e(Q_{Attack}, r_{Attack} s_{Attack} P)^{\upsilon'})$. If $H_3(e(Q_{Attack}, r_{Attack} s_{Attack} P)^{\upsilon'}) = H_3(e(Q_{B_1}, PK_{B_1})^{\upsilon}$ is tenable,

$\upsilon' = \upsilon$ must be satisfied. Since random number υ satisfies $\lambda_b = \upsilon P$, then its solution-finding means solving Elliptic Curve Discrete Logarithm Problem (ECDLP), which is computationally infeasible. Therefore, the probability $\mathrm{P}(\ell)$ of the attacker to decrypt the message c_1 is negligible. Consequently, the probability of attacker *A* winning the game is: $Adv(A) = |Pr[A]| = \frac{1}{2}\mathrm{P}(\ell)$, which is negligible. Therefore, the scheme realizes anonymity. However, since the optimal candidate is selected from k candidates, the probability of the attack inferring the optimal candidate is $1/k$. Let $\mathrm{P}(PID_k)$ be the probability of attacker solving the solution of d_i based on PID_k. So the probability that the attacker finally gets the privacy of the real requester is $\Pr(Sk) = \frac{1}{2k}P(\ell)P(PID_k)$. In the equation $PID_k = \boldsymbol{D}^{\boldsymbol{T}} \bullet \boldsymbol{d}_i$, D is an n-dimensional column vector randomly selected by TAS, so solving $\boldsymbol{d}_i$ on the basis of the equation $PID_k = \boldsymbol{D}^{\boldsymbol{T}} \bullet \boldsymbol{d}_i$ is infeasible. In addition, d_i is one of infinite solutions of the linear equation $\boldsymbol{\chi d_i = \rho}$, so the probability that the attacker guesses $\boldsymbol{d}_i$ is negligible. In summary, the probability of an attacker obtaining the privacy of the real requester is negligible.

5.3.2 Non-forgeability

Theorem 1 In Random Oracle Model (ROM), if attacker *A* exists to forge user's registration information by masquerading TAS in polynomial time, then Diffie-Hellman, the calculative problem, can be solved with non-negligible probability in polynomial time.

Proof: Assume attacker *A* is able to solve the calculative problem Diffie-Hellman with non-negligible probability in polynomial time, that is, attacker *A* finds $\boldsymbol{s}$ with non-negligible probability to make the equation $\mathrm{e}(S_U, P) \stackrel{?}{=} e(\tilde{Q}_U, PK_{anon})$ tenable.

Initialization: Assume challenger *C* possesses (P, sP) and provides for attacker *A* the system parameters $\{G_1, G_2, e, k, n, P, PK_{anon}, Q, H_1, H_2, H_3\}$, in which $PK_{anon} = sP$, while s is the random number generated by TAS , and is unknown to *C*; attacker *A* requests from *C* a random answer of ROM H_1 while maintains consistency to avoid conflict, and *C* keeps a request-reply list to store the replies from the requests.

ROM query phase: *C* is able to provide ROM query for attacker *A* via ROM H_1, and provide corresponding request-reply parameters.

Attacker *A* conducts query via ROM H_1 to obtain harsh values, as follows:

H_1 request: *A* requests the harsh value of identity ID_i from *C*, and *C* detects whether there is $ID_i \in L_I$ in request-reply list L_I;

(1) If there is $ID_i \in L_I$, then send the corresponding reply to *A*.

(2) Otherwise, randomly select $\tau_i \in Z_q^*$ and calculate $H_1(ID_i)$, send $(\tau_i, H_1(ID_i))$ to *A*, and store this request-reply in the list L_I, then the corresponding $S_{ID_i} = \tau_i H_1(ID_i)$ can be easily obtained.

Forgeability and problem-solving: attacker *A* forges user's registration information by masquerading TAS, but *A* is unable to obtain the random number s of TAS, so it fails to calculate S_U to make the equation $\mathrm{e}(S_U, P) \stackrel{?}{=} e(\tilde{Q}_U, PK_{anon})$ valid. If attacker *A* manages to obtain the random number s, then it has to speculate the random number s via the public key (P, PK_{anon}) and $PK_{anon} = sP$ in TAS, which means facing the calculative problem

Diffie-Hellman, so attacker *A* is unable to solve Diffie-Hellman problem with non-negligible probability in polynomial time. Therefore, the proposed scheme is able to meet the demand for non-forgeability.

5.3.3 Resistance to Query Tracing Attack

In the process of initiating LBS service, the initiator is replaced by the optimal candidate in launching LBS query, so in the query records of LBS server, what is recorded is the query information and ID information of the optimal candidate. Meanwhile, at different time intervals in users' mobility trajectories, there are different optimal candidates making continuous requests, so that attacker is unable to infer candidates' relevancy to real users via the intersection of user sets in cloak area at different time intervals. Therefore, it achieves the purpose of confusing the user's moving track. Assume the frequency of user's continuous query in mobility trajectories as m, and the number of candidates participating in each query is n_i, in which $1 \le i \le k$, since the optimal candidates in each query differ with one another, so the candidates in different cloak areas are independent with one another. Let $\mathrm{P}(PID_U)$ be the probability of attacker solving the solution of d_i based on PID_U, and $\mathrm{P}(\partial)$ be the probability of attacker capturing the communications between user U and candidate and decrypting the message, and if attacker obtains the registered message that MT user requests to TAS, then in the course of continuous queries the probability of tracing users is $\mathrm{P} = \prod_{i=1}^{m} \frac{1}{n_i} \mathrm{P}(\partial)\,\mathrm{P}(PID_U) \cdot \mathrm{P}(\partial)$ means solving the elliptic curve cryptosystem, which is infeasible in calculation. $\mathrm{P}(PID_U)$ is equal to the known $\boldsymbol{PID_u}$, then find the solution of d_i according to equation $PID_u = D^T \bullet d_i$, while D is a n-dimensional column vector randomly selected by TAS, and the probability of solving and finding d_i is negligible; moreover, there are infinite solutions of the linear equation $\boldsymbol{\chi d_i = \rho}$, so attacker is unable to identify $\boldsymbol{d_i}$ according to matrix equations. Therefore, it can be concluded that the probability of attacker obtaining requestor's real identity is negligible, that is, attacker is unable to trace candidates according to continuous query records so as to identify the real identity of the initiator.

5.3.4 Resistance to Wiretapping Attack

In this scheme, when user U communicate with TAS via MT, user *U* will employ TAS' public key PK_{anon} and LBS server's public key PK_{LBS} to encrypt message $m_1 = \{PID_C^1, PID_C^2, \cdots, PID_C^{k-1}\}$ and user's requested message $m_2 = \{L_U, Q_U, K_s\}$ so as to guarantee the security of messages during communications; and the optimal candidate B communicates with LBS server via secure channel. When returning query result, LBS server will adopts session key K_s to encrypt the query result $m_4 = \{MEG\}$ so as to guarantee its security. Finally, the optimal candidate B sends the encrypted data packet $M_{BoU} = \{En_{K_s}(m_4)\}$ to user U.

In conclusion, it can be demonstrated that the communications of LBS is conducted in encrypted manner to interchange messages, which effectively prevents the wireless communications among MT, TAS and LBS server from being monitored and wiretapped by attacker. Therefore, this scheme succeeds in resisting to wiretapping attack.

6. Simulation Experiment

The environment of simulation experiment in this scheme is as follows: operation system: Windows 7 (64 bit); CPU: Intel i5 processor; RAM: 4G; simulation software: mobile object generator Thomas Brinkhoff and simulation tool MATLAB. Suppose the experiment is conducted in an ideal network environment, then a large number of trajectory data of mobile objects are generated via Thomas Brinkhoff, and a certain mobile object is selected randomly as the initiator to realize confusion of users' trajectories by means of solution algorithm of trajectory similarity and query substitution algorithm, so as to further protect user's trajectory privacy. In comparison with this scheme, other schemes are adopted, like Random Scheme (based on fake trajectories randomly generated), schemes in Reference [11] and [16], as well as the optimal scheme in theory. The following experiment results are invariably the average value of 1000 operations.

As is shown in **Fig. 2**, given the same number of candidates K, the scheme in Reference [11] has the most unfavorable similarity, for it only considers user's marching direction after each pause in generating fake trajectories, while ignoring the map background information of fake trajectories; the scheme in Reference [16] has less unfavorable similarity, for it considers the properties of user's motion pattern and trajectory similarity in generating fake trajectories, and make reasonable adjustment according to locations' background information in fake trajectories, thus improving the trajectory similarity. In the proposed scheme of this paper, we adopts trajectory sampling and selects similar trajectories based on trajectory similarity function. Since this scheme conducts trajectory sampling at shorter time intervals, as a result, the candidate trajectories are quite short, which has a higher trajectory similarity than that of Reference [16].

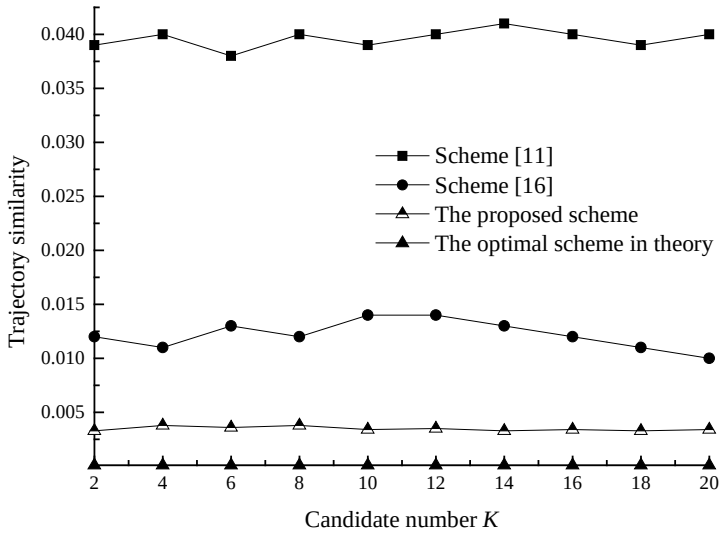

Fig. 2. The relationship between candidate number K and trajectory similarity

Given the fixed point number of trajectory segments, the processing time costs increase with increasing candidates, as is shown in **Fig. 3**. The explanation is: when increasing the candidate number of user or the point number within trajectory segments, certain time costs also increase because user need match more candidate users within certain range. Meanwhile, given the fixed candidate K, the execution time of algorithm increases with increasing point number N of trajectory segments. For example, when candidate K is 5, the point number N increases from 2 to 10, while the execution time increases from 19ms to 41ms, which means

the processing time is directly proportional to point number in trajectory segments. Consequently, the total time of searching the trajectory with the optimal similarity increases with increasing candidate user's number K and trajectory segments' point number N. It is to be noted that in real case this scheme must ensure that $K \geq 2$ and $N \geq 2$. When N=1, this trajectory is replaced by this sampling point, since a single point can not define the similarity of trajectory segments, so this scheme does not apply; if the environment in which the request users lie is excessively sparse, for instance, when $K = 0$, no candidate exists to substitute user in requesting service; when $K = 1$, that is, only one candidate exists, then there is no need to adopt trajectory similarity function in selecting optimal candidate, the only candidate is the optimal candidate to substitute user in requesting service, and attacker could trace it according to candidate's continuous queries and obtain the relevant private information of the initiator.

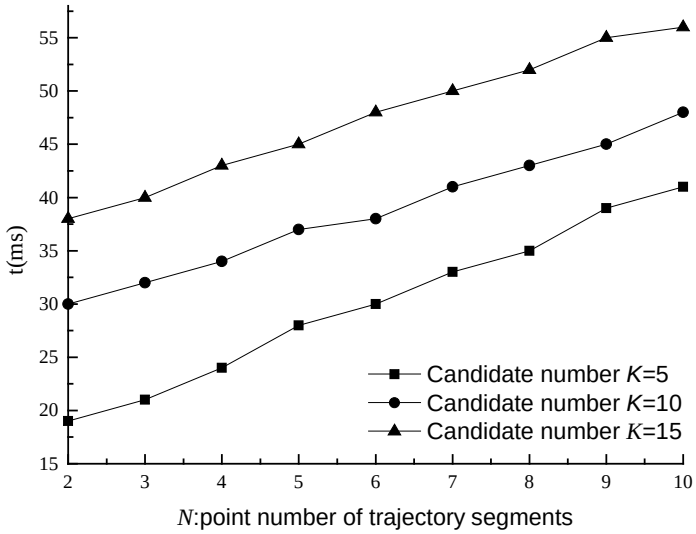

Fig. 3. The relationship among candidate number K, trajectory segments' point number N and the time t

In this scheme, given the fixed number N of trajectory segments, the influence of candidate K's variation upon the algorithm's execution time is shown in **Fig. 4**. According to the results, Random Scheme is the most efficient in execution for not taking into account other conditions, but is the most undesirable in protecting user's privacy via security analysis. The scheme in Reference [16] takes into account such elements as user's motion pattern, fake trajectories' validity and background information, etc., and requires more potential traversal and verification in generating and selecting fake trajectories, thus is much longer in algorithm's execution time. In this proposed scheme, what is needed is only the sampling of the existing trajectories, without generating trajectories, and selecting the trajectory with optimal similarity via trajectory similarity function, without considering other elements, therefore, the execution efficiency in this scheme is more desirable than that of Ref.[16] and Ransom scheme.

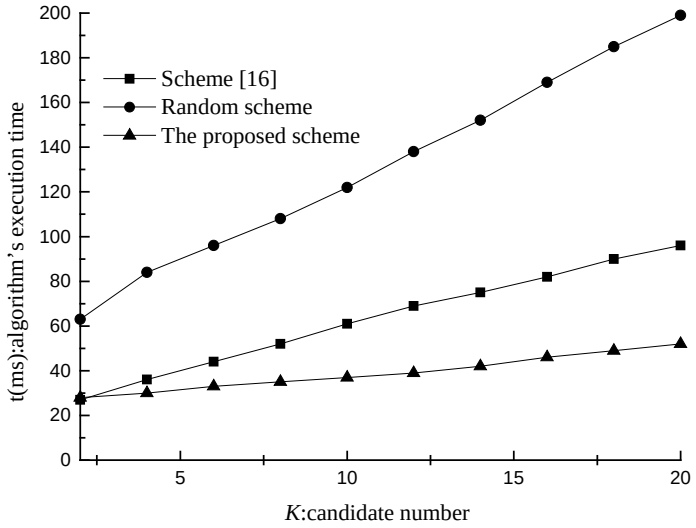

Fig. 4. The influence of candidate number K upon algorithm's execution time t

7. Conclusion

Aimed at the disclosure of mobile terminal user's location privacy in LBS, this paper proposes a trajectory privacy preservation scheme on the basis of similar trajectory's query substitution. Based on the anonymized identities of user and candidates, this scheme adopts trajectory similarity function to select the candidate with optimal trajectory similarity to substitute real user in requesting LBS service, so as to preserve user's privacy like identity, location trajectory and query contents. Security analyses prove that this scheme is not only able to solve such problems as anonymity and non-forgeability, but also able to resist continuous query attack and wiretapping attack. Moreover, simulation experiment is conducted in terms of three aspects: similarity between candidate and requested user, the influence of candidate number K and trajectory segments' point number N on time t, and the execution efficiency of the algorithm. The results show that the trajectory similarity between optimal candidate selected in this scheme and real user at certain time intervals is apparently more desirable than that of other schemes. Therefore, this scheme is of important theoretical significance and applicable value in mobile users' privacy protection in LBS.

Acknowledgements.

This work was supported by the National Natural Science Foundation of China (61872126, 61772159, 61300216); the Science and Technology Research Program of He Nan Province (192102210123, 182102110333, 172102310677).

References

[1] H. Huang, "Progress in Location-Based Services 2014," *Lecture Notes in Geoinformation & Cartography*, vol. 46, no. 6, pp. 0463, 2016. Article (CrossRef Link)

[2] Yanming.Suna , Min.Chena and L.Hu, "ASA: Against statistical attacks for privacy-aware users in Location Based Service," *Future Generation Computer Systems*, vol. 70, pp. 48-58, 2017. Article (CrossRef Link)

[3] W. He, "Research on LBS privacy protection technology in mobile social networks," in *Proc. of IEEE Advanced Information Technology, Electronic and Automation Control Conference, IEEE*, pp.73-76, 2017. Article (CrossRef Link)
[4] Ju X and Shin K G , "Location Privacy Protection for Smartphone Users Using Quad tree Entropy Maps," *Journal of Information Privacy & Security*, vol. 11, no. 2, pp. 62-79, 2015. Article (CrossRef Link)
[5] Naghizade E , Bailey J and Kulik L, "How private can i be among public users," *Acm International Joint Conference on Pervasive & Ubiquitous Computing, ACM*, pp. 1137-1141, 2015. Article (CrossRef Link)
[6] Zhang Yuan , Q. Chen and S. Zhong, "Privacy-preserving Data Aggregation in Mobile Phone Sensing," *IEEE Transactions on Information Forensics and Security*, vol.11, no. 5, pp. 980-992, 2016. Article (CrossRef Link)
[7] Wang H, Li FH and Niu B, "Advances in location privacy protection technology," *Journal on Communications*, vol. 37, no. 12, pp. 124-141, 2016.Article (CrossRef Link)
[8] Lee, Ken C. K., "Efficient Index-Based Approaches for Skyline Queries in Location-Based Applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 11, pp. 2507-2520, 2013. Article (CrossRef Link)
[9] Pan Xiao , X Meng and J Xu, "Distortion based Anonymity for Continuous Queries in Location Based Mobile Services," in *Proc. of Acm Sigspatial International Symposium on Advances in Geographic Information Systems DBLP*, pp. 256-265, 2009. Article (CrossRef Link)
[10] Freudiger Julien , R Shokri , and J. P. Hubaux, "On the Optimal Placement of Mix Zones," *Lecture Notes in Computer Science*, vol. 5672, pp.216-234, 2009. Article (CrossRef Link)
[11] Kato, R, Iwata, M and Hara, T, "A dummy-based anonymization method based on user trajectory with pauses," in *Proc. of International Conference on Advances in Geographic Information Systems* , pp. 249-258, 2012. Article (CrossRef Link)
[12] Zheng Huo, "PrivateCheckIn:Trajectory Privacy-Preserving for Check-In Services in MSNS," *Chinese Journal of Computers*, vol.36, no. 4, pp.716-726, 2013. Article (CrossRef Link)
[13] Ben Niu, Xiaoyan Zhu and Haotian Chi, "3PLUS: Privacy-preserving pseudo-location updating system in location-based services," in *Proc. of Wireless Communications & Networking Conference IEEE*, 2013. Article (CrossRef Link)
[14] Rui Chen, Benjamin C.M. Fung and Noman Mohammed, "Privacy-preserving trajectory data publishing by local suppression," *Information Sciences*, vol. 231, no. 1, pp. 83-97, 2013. Article (CrossRef Link)
[15] Zhao J, Zhang Y and Li Xing Hua, "A Trajectory Privacy Protection Approach via Trajectory Frequency Suppression," *Chinese Journal of Computers*, vol. 37, no. 10, pp. 2096-2106, 2014.
[16] LI Feng-hua, Zhang Cui and Niiu Ben, "Efficient trajectory privacy protection scheme," *Journal on Communications*, vol. 36, no. 12, pp. 114-123, 2015.
[17] Ye Ayong , Y Li and L Xu, "A novel location privacy-preserving scheme based on l-queries for continuous LBS," *Computer Communications*, vol. 98, pp. 1-10, 2016. Article (CrossRef Link)
[18] Sun Gang, Liao, Dan and Li Hui, "L2P2: A location-label based approach for privacy preserving in LBS," *Future Generation Computer Systems*, vol.74, pp.375-384, 2017. Article (CrossRef Link)
[19] X.H. Li, M.X. Miao and H. Liu, "An incentive mechanism for K -anonymity in LBS privacy protection based on credit mechanism," *Soft Computing*, vol. 21, no. 14, pp. 3907-3917, January 2017. Article (CrossRef Link)
[20] T. Lam and K. Rietsch, "Total positivity, Schubert positivity, and geometric Satake," *Journal of Algebra*, vol.460, pp.284-319, 2016. Article (CrossRef Link)

Cheng Song received his Ph.D. degree in Computer Science from the Beijing University of Posts and Telecommunications in 2011. He is now working as a lecture at the School of Computer Science and Technology in Henan Polytechnic University. His main research interests include Networking Security and Application, Privacy protection and Trusted Computing.

Yadong Zhang is a master student at the School of Computer Science and Technology in Henan Polytechnic University. His main research interests include Networking Security and Application, Privacy Protection.

Xinan Gu is a master student at the School of Computer Science and Technology in Henan Polytechnic University. His main research interests include Networking Security and Application, Privacy Protection and Anonymous Authentication.

Lei Wang received his Ph.D in control theory and engineering from Dalian University of Technology, China, in 2012. he is working as an associate professor in Henan Polytechnic University. His research interests include wireless Ad-hoc Networks, Embedded system, Networked control system, and Internet of things.

Zhizhong Liu received his PhD degree in Computer Science from Hohai University in 2011. He did his post-doc at Harbin Institute of Technology from 2013 to 2017. He is currently an associate professor in Henan Polytechnic University. He has authored or coauthored more than 30 papers. His research interests include Service-oriented Computing and Artificial Intelligence.