JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 36, 745-763 (2020)
DOI: 10.6688/JISE.202007_36(4).0004

Location Privacy Protection in Mobile Social Networks Based on *l*-diversity

HONG-TAO LI[1], LIN-XIA GONG[1], FENG GUO[2], QUAN-LI MIAO[3],
JIE WANG AND TAO ZHANG[4]
[1]*College of Mathematics and Computer Science*
Shanxi Normal University
Linfen, 041000 P.R. China
[2]*School of Information Science and Engineering*
Linyi University
Linyi, 276000 P.R. China
[3]*Veoneer China Co., Ltd.*
Shanghai, 201499 P.R. China
[4]*School of Computer Science and Technology*
Xidian University
Xi'an, 710071 P.R. China
E-mail: 25576152@qq.com

In recent years, location-based service has been widely used in social networks. However, people's locations or trajectory may be disclosed when they continuously use LBS to retrieve point of interests. The privacy disclosure problem not only restricts the development of LBS, but also reduces the quality of service. Recently, location privacy protection has attracted more and more attention. In this paper, aiming at dealing with the location privacy problem in mobile social network applications, we propose a location privacy protection method for multi-sensitive attributes based on *l-diversity* privacy protection model, and protect the user's location information in client side and server respectively. On the client side, the decomposition algorithm of minimum distance grouping is used to lighten the location data, which makes the processed data satisfy the l_1-*diversity* principle and upload the data to the server in the form of QIT^1 (Quasi-Identifier attribute Table) and ST^1 (Sensitive attribute Table) to achieve the initial protection of the user's location data. On the server side, the minimum selection priority strategy is adopted to form the l_2-*diversity* group satisfying the multi-sensitive attributes, and the data is uploaded in the form of QIT^2 and ST^2 to further protect the user location data (where $l_1 < l_2$). The experimental results show that this method not only can effectively protect location privacy data, but also has high data availability.

Keywords: location based services, mobile social network, location privacy, *l-diversity*, privacy protection

1. INTRODUCTION

With the continuous development of mobile Internet and mobile terminal technology, social network has been widely used in people's lives, and information dissemination becomes more efficient via the social network [1]. In social networks, users can use mobile terminal device to send their demands to mobile social network server to obtain relevant services. This application mode provides users with a social platform to share information such as interests, hobbies, activities and status in social communication network. At pre-

Received September 10, 2019; revised November 7 & December 6, 2019; accepted December 9, 2019.
Communicated by Xiaohong Jiang.

sent, there are many Social Network Services, check-in applications [2], and other applications that share location information content on social networks [3]. These application services facilitate the sharing of information among users. However, when users share location information, it is likely to cause privacy disclosure. With the development of social networks and GPS, Location Based Services (LBS) application [4] has become a mainstream application in mobile social network applications.

Location information of the mobile terminal is obtained through radio communication network or external positioning method (such as GPS), and provides many convenient services for the user. However, location service information usually contains user's private information, such as user's life track, personal habits and social situations. If it is not handled properly in the process of data collection and transmission, users will face various risks of privacy disclosure, which threatens the privacy information [5-9]. Therefore, protecting the security of users' location privacy information is a research hotspot in privacy protection [10-13].

In recent years, domestic and foreign scholars have paid more and more attention to privacy and security of data [14]. Machanavajjhala *et al.* [15] proposed the concept of *l-diversity*, in which the diversity of sensitive attributes in each anonymous attribute group satisfies greater than or equal to *l*, effectively ensuring the privacy protection of single-sensitive attribute publishing data sets. Sun *et al.* [16] proposed an extended (l, α)-*diversity* model which satisfies *l-diversity* and requires that the total weight of sensitive values in each equivalence group should not be lower than the threshold value α. Most existing data privacy protection technologies [17-19] have been only applicable to the privacy protection of single-sensitive attribute data, but not to the privacy protection of multi-sensitive attribute data. Yang *et al.* [20] studied the issue of multi-sensitive attributes privacy data release in detail, and proposed a multi-dimensional bucket grouping technology based on loss connection, which was applicable to the security issue of privacy data with multi-sensitive attributes. In the privacy protection method for multi-dimensional sensitive attribute data publishing, the personalized privacy problem of attribute value weights of numerically sensitive attributes is rarely considered. Lu [21] proposed a personalized privacy protection method based on clustering and weighted multi-dimensional bucket grouping. This method divides the attribute values of each dimension numerically sensitive attribute into multiple clusters by clustering, and constructs a weighted multi-dimensional bucket by multi-dimensional numerically sensitive attributes. The group satisfying the *l-diversity* is formed by the principle of the maximum dimension capacity priority of the weighted selectivity, and the group is obtained to be published in the form of an anonymous table. Zhu *et al.* [22] proposed an *l-diversity* algorithm based on segmentation and clustering, which used variance to calculate the weight of data attributes, calculated the comprehensive value of each record, divided the equivalence class according to the comprehensive value, and then realized the continuous release of data. Han *et al.* [23] proposed a hierarchical *l-diversity* model for numerically sensitive attributes, including hierarchical disparity *l-diversity*, hierarchical information entropy *l-diversity*, and hierarchical recursion (*c*, *l*)-*diversity*. Firstly, the model classifies the numerical sensitive attribute fields, then realizes the *l-diversity* of the numerical sensitive attributes based on the hierarchical information, and implements the *l-Incognito* algorithm of the model. The above literatures propose various solutions for the privacy protection of data and location information. However, due to the particularity of mobile social network applications, users often share the convenience of location information exchange in order to obtain better services, and loca-

tion information is transmitted through the mobile Internet, which brings new challenges to the protection of location privacy.

This paper mainly deals with the problem of privacy location disclosure of users in mobile social network applications. When the quasi-identifier attribute dimension of the user's location data table is large, if the attribute dimension reaches the exponential level, the *k-anonymity* protection of the location data can be directly processed by generalizing and anonymous methods, which will lose a lot of location data and affect the availability of the data. In this paper, the location data table is divided into the quasi-identifier attribute table (*QIT*) and the sensitive attribute table (*ST*) respectively. The latitude and longitude of the user's location can be regarded as two sensitive properties, and the privacy location data can be protected through loss-connection. Based on this idea, *l-diversity* protection method is more appropriate.

The contributions of this paper are as follows. Firstly, a location privacy protection system architecture and its threat model are proposed, and various security issues therein are illustrated. Secondly, an *l-diversity* based location privacy protection method in mobile social network applications is proposed, which performs location information protection processing at the client and server respectively. On the client side, the original data is pre-processed by lightweight decomposition algorithm of minimum distance grouping to make data satisfy l_1-*diversity* principle. On the server side, each tuple in the location data table is formed into a multi-sensitive attribute data group satisfying the l_2-*diversity* principle by adopting the minimum selection priority strategy, and the resulting groups are uploaded in the form of QIT^2 and ST^2, realizing further protection of location data (where $l_1 < l_2$). Finally, we conduct detailed theory analysis and a comprehensive set of experiments to show our method is effective for privacy protection with low information loss and computing time.

The rest of the paper is organized as follows. Section 2 of this paper introduces the system architecture and threat model based on *l-diversity* privacy protection method. Section 3 introduces the definition and implementation process of location privacy protection method based on *l-diversity* in mobile social networks; Section 4 is Experimental results and analysis; Section 5 summarizes the contents of this paper.

2. SYSTEM ARCHITECTURE AND THREAT MODEL

2.1 System Architecture

The system architecture of this paper is shown in Fig. 1. The whole system architecture is mainly composed of client, privacy protection processor and location service provider. The client consists of three modules: GPS positioning module, raw database and lightweight processor, which are mainly responsible for storing and light weighting the user's location data, and storing the personalized privacy protection parameters l_1 and l_2 (where $l_1 < l_2$), a predefined set of sensitive semantic locations and sensitivity level classification; The privacy protection processor consists of three modules: location protection module, pre-processed database and query processing module. The privacy protection model is mainly responsible for further processing the location data of the user, and the query processing module achieves the protection of the user query data through the related query method; the location service provider is not trusted, users can make relevant query requests to them and get their corresponding feedback on location information.

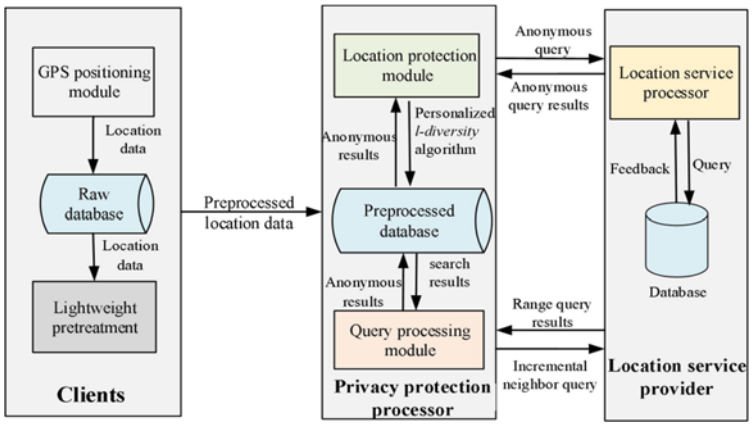


Fig. 1. Location privacy protection system architecture.

Aiming at the problem of location privacy disclosure of users in mobile social networks, this paper proposes a privacy protection method of multi-sensitive attributes based on *l-diversity*, which protects the location privacy information of users on the client and server respectively. On the client side, the user uses the GPS module of the mobile social network application to locate the location and upload the data to the raw database. Then, the preprocessing module uses the decomposition algorithm of minimum distance grouping to lightweight preprocess the raw data, so that the processed data satisfies the l_1-*diversity* principle and uploads the data to the privacy protection processor in the form of QIT^1 and ST^1 to avoid the attacker's attack on the raw data. Assuming that the privacy protection processor as the trusted third party, on the location protection processor, the location protection module obtains the location data in the pre-processed database and processes it with the minimum selection priority algorithm, which makes the processed data satisfies the multi-sensitive attributes l_2-*diversity* principle, and uploads the obtained groups in the form of QIT^2 and ST^2 to realize further protection of the location data. When the location protection module puts forward an anonymous query to the location service provider, database in the location service provider makes a query request to the location service processing module, then the location service processing module feeds back the query request of the database, and finally feeds back the query result to the location protection module and uploaded to the preprocessed database.

2.2 A Practical Application Scenario

A practical scenario illustrates in Fig. 2. is the social network application in smartphone, which brings people a lot of convenience in the life. Our intention is to protect and process location information of social applications in smart phones. The client refers to APPs in mobile phones in social network, and the server refers to Location service providers. Server provides APPs API interface to obtain the relevant data, adopt the minimum distance grouping algorithm to protect the data, upload the processed data to APPs database in the privacy protection processor, adopt the minimum selectivity priority algorithm to achieve secondary protection of the location data, finally upload the processed data to the database. Users can obtain location query results from the APPs service providers.

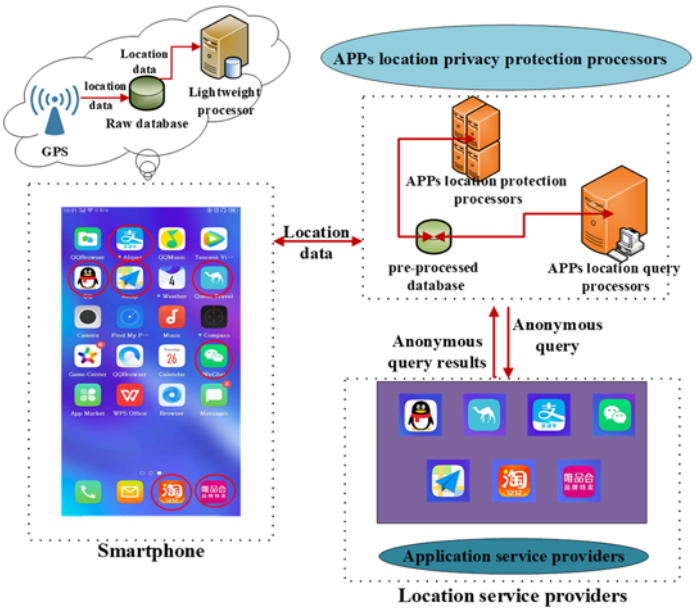


Fig. 2. A practical application scenario in Smartphone.

2.3 Threat Model

This paper assumes that attackers attack users' sensitive location information periodically and passively. There are many service providers with different security guarantees at present, such as Google, Baidu, Alibaba. If these service providers are attacked, sensitive information of users will be disclosed. Therefore, this paper assumes that an attacker can make periodical and passive attack on location service provider in network. Based on the assumption, the privacy location information of the user is not protected, once the attack happens, the privacy information of the user will be disclosed. The threat model of location privacy system is shown in Fig. 3.

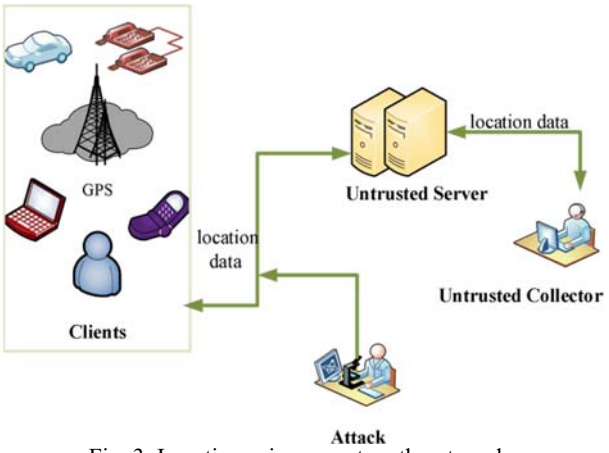


Fig. 3. Location privacy system threat mode.

Aiming at the above problems, this paper proposes a multi-sensitive attribute privacy protection method based on *l-diversity*. On the client side, lightweight preprocessing is carried out to ensure the user's location privacy is initially protected. Untrusted location service providers will provide location data to attackers for commercial purposes, resulting in the disclosure of users' location privacy. Therefore, further protection of location data is required between the user and the service provider.

3. LOCATION PRIVACY PROTECTION METHOD IN MOBILE SOCIAL NETWORKS BASED ON *L-DIVERSITY*

3.1 Preliminary Knowledge

In mobile social network applications, users upload location information through mobile devices with location functions. Location information is divided into accurate location information (such as *GPS* coordinates) and semantic location information (such as *POI* information). In this paper, g_i is used to represent location information (where $1 \le i \le n$), $G = \{g_1, g_2, \ldots, g_n\}$ is the set of location g_i, and represent the location space of mobile social network. User-uploaded location information can be expressed as $g_i = (u_i, t_i, la_x, lo_y)$. Where, u_{id} represents the identity attribute of the user, (lo_x, la_y) represents the coordinate attribute of the longitude and latitude of the user's geographical location, and t_i represents the times tamp attribute for location information. The format of common location data is shown in Table 1.

Definition 3.1: *Quasi-Identifier Attributes*: A set of attributes in a data table $T(id, q_1, \ldots, q_m, s)$ that can be used to identify a user by linking to an external data table. For example, the *quasi-identifier attributes is* $\{u_{id}, t_i\}$ in Table 1.

Definition 3.2: *Sensitive Attributes*: Attributes that need to be strongly protected during data upload or publication. For example, the sensitive attribute $S = \{lo_x, la_y\}$ in Table 1.

Definition 3.3: *Equivalence Class*: A set of tuples in data table *T*, where each tuple has the same attribute value of quasi-identifier.

Table 1. The raw data table.

Location number	u_{id}	t_i	lo_x	la_y
g_1	02147	t_1	111.513	36.096
g_2	02178	t_2	111.514	36.086
g_3	11456	t_3	125.511	45.088
g_4	11435	t_4	142.401	72.918
g_5	47531	t_5	130.301	51.813
g_6	47501	t_6	140.498	48.426

Definition 3.4: *L-Diversity*: When an anonymous data table satisfies *k-anonymity*, the same equivalence class contains at least one different sensitive attribute value, so that the probability that the sensitive attribute being attacked is at least. For example, Table 2 is an anonymous information table that satisfies 2-*diversity*.

Table 2. 2-diversity anonymous information table.

Location number	u_{id}	t_i	lo_x	la_y
g_1	021**	$\{t_1, t_2\}$	111.513	36.096
g_2	021**	$\{t_1, t_2\}$	111.514	36.086
g_3	114**	$\{t_3, t_4\}$	125.511	45.088
g_4	114**	$\{t_3, t_4\}$	142.401	72.918
g_5	475**	$\{t_5, t_6\}$	130.301	51.813
g_6	475**	$\{t_5, t_6\}$	140.498	48.426

Definition 3.5: *Selectivity of the Tuple*: The selectivity $Select(g_i)$ of tuple g_i is the sum of the frequencies of the sensitive attribute value s_v of the longitude and latitude in g_i within the range of sensitive semantic location preset by the user, that is

$$Select(g_i) = \sum_{s_v \in s_{set}(g_i)} f(s_v). \tag{1}$$

Where, $s_{set}(g)$ is the set of all longitude and latitude sensitive attribute values in the tuple g_i, and $f(s_v)$ is the frequency at which the longitude and latitude sensitive attribute value s_v appears within the range of sensitive semantic location preset by the user.

Definition 3.6: *Personalized Selection Degree of Tuple*: The personalized selection degree $PSelect(g_i)$ of tuple g_i is the sum of the frequencies of sensitive attribute values of different sensitivity degrees in g_i within the range of sensitive semantic location preset by the user, that is

$$PSelect(g_i) = \sum_{s_V \in s_{set}(g_i)} f(s_v) \times TDegree(g_i). \tag{2}$$

Where, $TDegree(g_i)$ is the sensitivity of tuple g_i.

In order to solve the problem of location information disclosure of mobile social networks, this paper proposes a multi-sensitive attribute location privacy protection method based on *l-diversity*, which protects users' location information on the client and server respectively. On the client side, the decomposition algorithm of minimum distance grouping is adopted to carry out lightweight preprocessing for the location data, so that the processed data satisfy the l_1-*diversity* principle and upload the data to the server in the form of QIT^1 and ST^1 to realize the initial protection of the user's location data. On the server side, the minimum selection priority strategy is adopted to form the l_2-*diversity* group satisfying the multi-sensitive attribute, and the data is uploaded in the form of QIT^2 and ST^2 to further protect the user location data (where $l_1 < l_2$).

3.2 Client Minimum Distance Grouping Algorithm

On the client side, the accurate location information is regarded as two sensitive attributes: longitude sensitive attribute lo_x and latitude sensitive attribute la_y. A minimum distance grouping decomposition algorithm is proposed based on different values of multidimensional sensitive attributes. Firstly, the longitude sensitive attributes lo_x and latitude sensitive attributes la_y of location information are sorted according to the size of sensitive

attribute values. Then the benchmark data is selected and the optimal tuple is selected to join the current group each time according to the idea of minimum distance grouping, and the group satisfies the l_1-*diversity* principle. Then, it checks whether the remaining tuples can be added to an equivalent class one by one without destroying the l_1-*diversity* principle of the equivalent class. If it is satisfied, it is added to the equivalence class, otherwise the tuple is hidden. Finally, the resulting groups are uploaded in the form of QIT^1 and ST^1 to realize the initial protection of the user location data. In this paper, the minimum value of each dimension of sensitive attributes is taken as the benchmark data, which is recorded as $s_{\min} = \{lo_{x\min}, la_{y\min}\}$. The Euclidean distance is used to measure the length of two sensitive attribute vectors, and the tuple of the minimum distance $s_{\min}$ is added to the grouping. The per-dimensional sensitive attribute value on s is close to $s_{\min}$.

$$\rho = \sqrt{\left(lo_{x\min} - lo_{xi}\right)^2} + \sqrt{\left(la_{y\min} - la_{yi}\right)^2} \tag{3}$$

Where, ρ represents the length of the benchmark data $s_{\min}$ and the two sensitive attribute vectors in any unary group in the data table.

The details of minimum distance grouping decomposition algorithm are described as follows.

Algorithm 1. Minimum distance grouping algorithm

Input: Location data table T, Parameter l_1

Output: Quasi-identifier table QIT^1, Sensitive attribute table ST^1

1. $vec[lo_{xn}] \leftarrow [lo_{xmin}]$; //Sort by longitude value
2. $vec[la_{xn}] \leftarrow [la_{xmin}]$; //Sort by dimension value
3. $s_{min}(lo_{xmin}, la_{xmin})$;

//The tuple with the smallest sensitive attribute value is selected as the benchmark data

//Group stage

4. Group $EC = \phi$ on data table T
5. $i = 0$
6. $\rho\ (s_{min}, g_i)$

//Calculate the distance between the benchmark data and other tuples

7. **while** $(n \% l_1) > i$ **do**
8. $EC_i = l_1 + 1$
9. **if** $EC_i = l_1$ **then**
10. add the record $vec[i * (n / l_1)]$ to EC_i
11. $i++$
12. Delete the record g_i added to EC_i from the data table T
13. **else** $l_1 >= 2$ **then**
14. anonymous failure
15. **end if**
16. **end while**

//Processing the remaining tuple stage

17. **while** $g_i > 0$ //There are still remaining tuples in T
18. **if** adding g_i to EC_i still satisfies l_1-diversity **then**
19. i + 1 //Add one to the record in group EC_i
20. **else** hide these records
21. **end while**
22. **exit**
23. **Output** all group EC_i in the form of QIT^1 and ST^1

Fig. 4. Minimum distance grouping algorithm.

Firstly, steps 1-2 arrange values of each sensitive attribute from small to large; Secondly, The tuple with the smallest sensitive attribute value in each dimension is selected as the benchmark data (step 3); Next, grouping stage (steps 4-18), grouping initialization Selects the in/l_1 tuples in turn and calculates the distance between benchmark data and other tuples based on Euclidean distance, and select l_1 (when $n\%l_1 = 0$) or $l_1 + 1$ (When $n\%l_1 != 0$) (steps 4-6). The tuples are grouped as an equivalence class, and the above process is repeated until the requirement to constitute an equivalence class group cannot be satisfied (steps 7-16); Then, perform concealment processing on the remaining data tuples one by one to check whether the remaining data tuples can be added to an equivalent class group EC_i. If the l_1-*diversity* principle is satisfied, the tuples can be added to the equivalent class EC_i. Otherwise, the tuple can be hidden, as so the hiding rate of the data uploading is increased. In this way, the selected group on each dimension sensitive properties can not only satisfy approximately l_1-*diversity* but also minimize the difference in the group in each dimension of sensitive attributes (steps 17-21); Finally, the processed location data is grouped and uploaded in the form of QIT^1 and ST^1 (step 22).

Taking the location data in Table 1 as an example, the algorithm execution process is described. Adopting $l = 2$ as the parameter grouping, the minimum distance grouping algorithm first sorts each dimensional sensitive attributes of the location data according to the longitude and latitude values to obtain the longitude $\{g_1, g_2, g_3, g_5, g_6, g_4\}$ and the latitude $\{g_2, g_1, g_3, g_6, g_5, g_4\}$, and selects the tuple g_1 with the smallest longitude dimension and latitude dimension. The tuple g_2 with the smallest value is recorded as $s_{\min} = \{g_1 longitude, g_2 latitude\} = \{111.513512, 36.086712\}$, and the tuple with the smallest distance from $s_{\min}$ is g_2, which is added to EC_1. The $n/l = 6/2 = 3$ tuple is g_5, g_6, recorded as $s_{\min} = \{g_5 longitude, g_6 latitude\} = \{130.301529, 48.426926\}$, and the tuple with the smallest distance is g_5 and g_5 is added to the EC_1, a group is obtained as $\{g_2, g_5\}$. These tuples are deleted in the vector, and the algorithm loops. Finally, the upload location data is shown in Table 3.

Table 3. Minimum distance grouping algorithm result.

	Location number	u_{id}	t_i	Group
QIT	g_1	021**	$\{t_1, t_2\}$	EC_2
	g_2	021**	$\{t_1, t_2\}$	EC_1
	g_3	114**	$\{t_3, t_4\}$	EC_3
	g_4	114**	$\{t_3, t_4\}$	EC_3
	g_5	475**	$\{t_5, t_6\}$	EC_1
	g_6	475**	$\{t_5, t_6\}$	EC_2

	Group	Sensitive Attributes
ST	EC_1	(111.514, 36.086) (130.301, 51.813)
	EC_2	(111.513, 36.096) (140.498, 48.426)
	EC_3	(125.511, 45.088) (142.401, 72.918)

3.3 Server Side Minimum Selection Priority Algorithm

On the server side, an algorithm based on *l-diversity* with minimum selectivity priority is proposed. The minimum selectivity priority strategy is adopted to group the tuples in the location data table, forming a group that satisfies the multi-sensitive attribute l_2-*diversity* principle. Then, it checks whether the remaining tuples can be added to an equivalent class of EC_i and satisfy the multi-sensitive attribute l_2-*diversity* principle of the equivalent class. If the principle is satisfied, the tuples can be added to the equivalent class of EC_i, otherwise, the tuples can be hidden. Finally, the processed data groups are uploaded in the form of QIT^2 and ST^2, which achieves further protection of location data.

Different sensitive attributes have different sensitive values. Users can predefine the sensitivity values of longitude and latitude sensitive attributes and predefine the sensitive semantic locations in the location data table, and then classify them according to the sensitivity of different sensitive values. The exact locations belonging to the same semantic location range are regarded as the same sensitivity level. The $SDegree(S_i)$ is set to represent the sensitivity of S_i grade, and the weight $\omega_{i,i\text{-}1}$(where $2 \le i \le d$) is used to represent the connected weight between the adjacent levels of S_i, Where the sensitivity of the sensitive value is equal to that of the group to which the sensitive value belongs. The initial condition is $SDegree(S_1) = 0$, $SDegree(S_d) = 1$, Where the conditional relationship of the sensitivity between the adjacent levels of S needs to be satisfied as:

$$\frac{SDegree(S_{i+1}) - SDegree(S_i)}{SDegree(S_i) - SDegree(S_{i-1})} = \frac{\omega_{i+1,i}}{\omega_{i,i-1}}, 2 \le i \le d. \tag{4}$$

There are two different definitions of the value of $\omega_{i,i\text{-}1}$:

(1) $\omega_{i,i\text{-}1} = 1 (2 \le i \le d)$. The connected weight between all adjacent levels is 1.

(2) $\omega_{i,i-1} = \frac{1}{(i-1)^{\beta}} (2 \le i \le d, \beta \ge 1)$. It's fixed for β.

This paper assumes that all adjacent levels have a connection weight of 1. In the process of uploading location data, the sensitivity level is divided according to the sensitivity of sensitive values of longitude and latitude sensitive attributes, and then the sensitivity of each level S_i is calculated by the sensitivity level relationship between adjacent levels. The sensitivity $TDegree(g_i)$ of the tuple is the sensitivity of the sensitivity level to which the sensitivity value of longitude and latitude sensitive attributes belongs. That is

$$TDegree(g_i) = SDegree(g_i). \tag{5}$$

The detail of minimum selection priority algorithm is described as Fig. 5. Firstly, steps 1-3 calculate the sensitivity and personalize selectivity of each tuple in the location data table T; Secondly, steps 4-7 rank the size of the personalized selectivity of each tuple, initialize the tuple group, select the tuple with the minimum personalized selectivity as the initial class member of the equivalence class. In the grouping stage, the tuples of sensitive values with different sensitive attributes are put into their equivalence classes to form an equivalence class group that satisfies the principle of location-sensitive attribute l_2-*diversity*, and the tuples with the same sensitive value are placed in the set to be grouped and the above process is cyclically executed until the requirements for forming an equivalence class group cannot be satisfied (steps 8-15); Then, the remaining tuples are processed by adding to the group of an equivalent class in turn. If the l_2-*diversity* principle of location-sensitive attributes of the equivalent class is satisfied, add it into this equivalence class, otherwise hide it (steps 15-20). Finally, the resulting groups are uploaded in the form of QIT^2 and ST^2 (step 21).

Algorithm 2. Minimum selection priority algorithm

Input: Location data table T, Parameter l_2, Sensitivity classification table $T_1 \ldots T_d$
Output: Quasi-identifier attribute table QIT^2,Sensitive attribute table ST^2

1. $TDegree(g_i) = SDegree(g_i)$ //The sensitivity of each tuple is calculated
2. $Select(g_i) = \sum_{s_v \in s_{\_}(g_i)} f(s_v)$
3. //Calculating the sum of the frequencies at which the sensitive attribute value s_v appears within the range of sensitive semantic locations preset by the user
4. $PSelect(g_i) = \sum_{s_v \in s_{\_}(g_i)} f(s_v) \times TDegree(g_i)$ //Calculate the personalized selectivity of tuples in T
5. vec[n] ← [min] //Sort by PSelect(g_i)

//Group stage

6. $EC = \emptyset$
7. i = 0
8. **While** $EC_i \leftarrow g_{min}$ **do** //Add PSelect(g_i) minimally to the equivalence class grouping
9. $g_{min} = g_i$
10. **if** $EC_i \leftarrow g_i$ **then**
11. i + + //Increased tuples in EC_i
12. The record g_i added to the EC_i is deleted from the data table T
13. **else** i >= l_2 **then**
14. EC_i to be uploaded
15. **end if**
16. **end while**

//Processing the remaining tuple stage

17. **while** $g_i > 0$ //There are still records remaining in T
18. **if** after adding g_i to EC_i, it still satisfies l_2-diversity
19. i + 1 //Add one to the record in the group EC_i
20. **else** hide these tuples
21. **end while**
22. **exit**
23. **Output** all group EC in the form of QIT^2 and ST^2

Fig. 5. Minimum selection priority algorithm.

3.4 Algorithm Analysis

(A) Safety Analysis

This paper protects the location privacy of users on the client side and server side respectively based on the *l-diversity* anonymous model. On the client side, the minimal distance grouping decomposition algorithm is used to generate the user's location information into n/l groups, where l (when $n\%l = 0$) or $l+1$ (when $n\%l! = 0$) the tuples form a group. The greater the minimum difference within each group in each dimension of sensitive attributes, the higher the security of uploaded results is, making it impossible for an attacker to approximate the privacy location. Since the location data is uploaded in the form of a group, the probability of the attacker inferring the privacy location in-formation is about $\frac{1}{n^2}$. On the server side, the location data of users are generated into n/l groups by using the strategy of minimum selection priority. In each equivalence class, there are at least l tuples, and the similarity between the tuples is extremely small. Therefore, the probability that an attacker can speculate the precise location of the user is about $\frac{1}{n^2}$, and the probability that an attacker can speculate the precise location of the user is $\frac{1}{n^2} \times \frac{1}{n^2} = \frac{1}{n^4}$.

(B) Complexity Analysis

Assuming that the location data table contains n records data, in the process of using the minimum distance grouping algorithm on the client, the size of the attribute values of the latitude and longitude sensitive attributes in the location data table are respectively sorted, and the time complexity is $O(n^2)$. The time complexity of calculation of the distance between benchmark data and other tuples is $O(n\log n)$. The time complexity of sorting the euclidean distance is $O(n^2)$. In the grouping phase, the time complexity is $O(n)$. The worst time complexity of the processing stage of the remaining tuple is $O(n)$, so the time complexity of using the algorithm on the client is $O(n^2)$. In the process of the minimum selection priority algorithm on the server side, the personalized selectivity of each tuple is calculated, and the time complexity is $O(n)$; The time complexity of sorting personalized selectivity is $O(n^2)$. Only one selection judgment is performed for each tuple in the tuple classification stage, so the time complexity of the classification stage is $O(n)$; The processing stage of the remaining tuples is also the process of selecting and judging the remaining tuples. In the worst case, the time complexity is $O(n)$. In summary, the overall complexity of the algorithm is $O(n)+O(n^2)+O(n)+O(n^2) \approx O(n^2)$.

4. EXPERIMENTAL RESULTS AND ANALYSIS

To test the proposed location privacy protection method, it runs on the hardware environment of *Intel*(*R*) *Core*(*TM*) *i*5-4210*M CPU*@2.6*GHz* processor and 4*GB* memory Windows 7 64-*bit* operating system. The experimental data set is real location data from the mobile social network *Gowalla* [24], the data set collected 1679245 POI records visited by social friend relationships among 329839 users from September to December 2011, as well as the classification of location types visited.

The experiment mainly compares and analyzes the performance indexes of the algorithm in three aspects: (1) the data sets $|T|$ of different sizes; (2) the values of different

diversity parameters l; (3) the dimension of different sensitive attribute d. First, we test the information loss and anonymity ratio of two kinds of privacy data upload algorithms: the minimum distance grouping algorithm and the minimum selection priority algorithm through a large number of experiments, and compare them with the information loss and anonymity ratio of the maximum bucket priority algorithm [20] of privacy data release of multi-sensitive attributes, and comprehensively analyze the advantages and disadvantages of the two algorithms in the aspects of information loss and privacy protection of upload data.

4.1 Anonymity Ratio Analysis

The experiments test the influence of different diversity parameter l values, different data sets $|T|$ and different sensitive attribute dimensions d on anonymity ratio as shown in Fig. 6. We set the dimension of the sensitive attributes $d=2$, $l=3$. The anonymity ratio of the minimum selection priority algorithm and the minimum distance grouping algorithm is less than of the maximum bucket priority algorithm. This is because the number of buckets in the maximum bucket priority algorithm is smaller than in the minimum distance grouping algorithm and the minimum selection priority algorithm, so its grouping diversity selection is more, and the grouping success rate is also greater. The anonymity rate of the three algorithms decreases with the increase of data set $|T|$ as shown in Fig. 6. This is because with the increase of data set, the number of tuples satisfying *l-diversity* principle increases, which resulting in a lower anonymity rate. The anonymity rate of the minimum selection priority algorithm does not exceed 1.8% for different data sets, which indicates the algorithm has good performance and data availability.

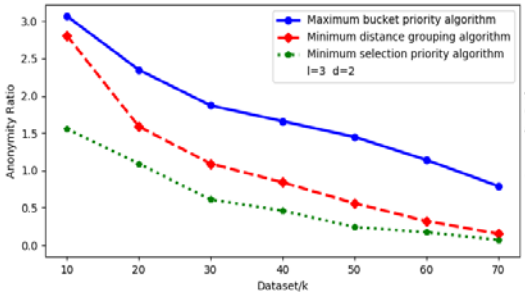


Fig. 6. Comparison of anonymity ratio under different data sets.

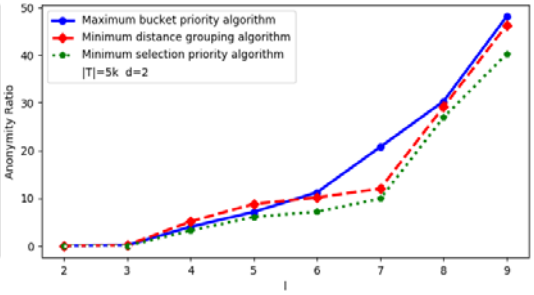


Fig. 7. Comparison of anonymity ratio under different l values.

In Fig. 7, we set the dimension of sensitive attributes $d=2$ and the data set $|T| = 5k$. The anonymity rate of the minimum selection priority algorithm and the minimum distance grouping algorithm is less than the maximum bucket priority algorithm. As the value of l increases, the more different sensitive values are included in the generated equivalence class group, the higher requirement for diversity, which can lead to the hidden tuple more. In Fig. 8, we set the data set $|T| = 5k$ and $l=3$, the anonymity rate of the three algorithms increases with the dimension of sensitive attributes. This is because the more the dimension

of sensitive attributes, it is more difficult for the group to satisfy the *l-diversity* principle, which resulting in the increase of the number of tuples requiring to be hidden.

As can be seen in Fig. 8, with the increase of the dimension of sensitive attributes, the anonymity rate of the minimum selection priority algorithm is less than that the minimum distance grouping algorithm and the maximum bucket priority algorithm. When the number of sensitive attributes $d = 2$, the anonymity rate of the three algorithms is close to 0, and the optimal grouping result is obtained. Compared with the maximum bucket priority algorithm, the minimum distance grouping algorithm and the minimum selection priority algorithm can always ensure that each group satisfies the *l-diversity*.

4.2 Information Loss Analysis

The experiment analyzes the influence of different diversity parameter l values, different data sets $|T|$ and different sensitive attribute dimensions d on the information loss of different algorithms. In Fig. 9, we set the dimension of the sensitive attributes $d = 2$, $l = 3$. For different data sets, the information loss of the three algorithms decreases with the data sizes. This is because as the data sizes increases, the number of tuples that satisfy the *l-diversity* principle increases, the fewer tuples that need to be hidden, the lower the amount of information loss. As the data sizes increase, the information loss of the three algorithms is close to 0.

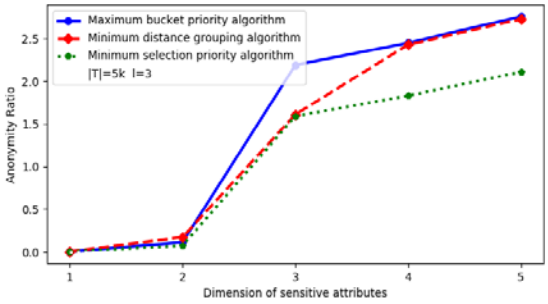


Fig. 8. Comparison of anonymity ratio under different dimensions of sensitive attributes.

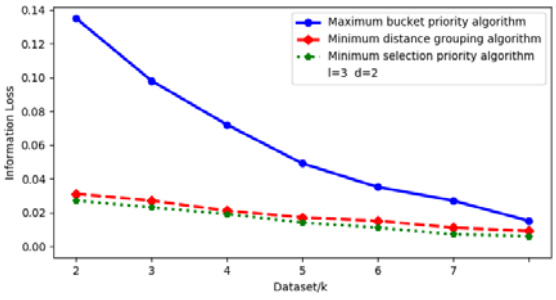


Fig. 9. Comparison of information loss under different data sets.

In Fig. 10, we set the dimension of sensitive attributes $d = 2$ and the data set $|T| = 5k$. The information loss of the three algorithms increases with the increase of l value, when $l = 2$, the information loss of the three algorithms did not exceed 0.03, but the information loss of the three algorithms increased accordingly, this is because as the value of l increases, it is more difficult to make the group satisfy the *l-diversity* principle, resulting in more information loss.

As shown in Fig. 11, we set the data set $|T| = 5k$ and $l = 3$, the information loss of the three algorithms increases with the increase of the dimension of the sensitive attributes, and compared with the other two algorithms, the information loss of the minimum selection priority algorithm is lower than that of the other two algorithms. This is because the increase of the dimension of the sensitive attributes limits the success rate of the anonymity of the group and also causes the information loss to become larger and larger.

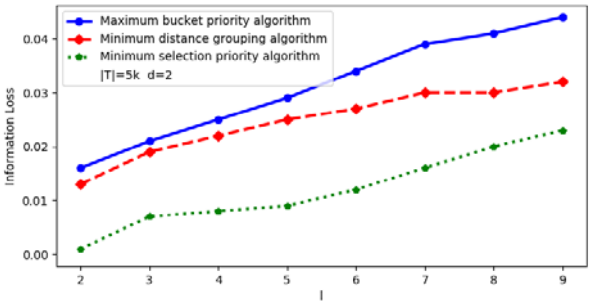


Fig. 10. Comparison of information loss under different l values.

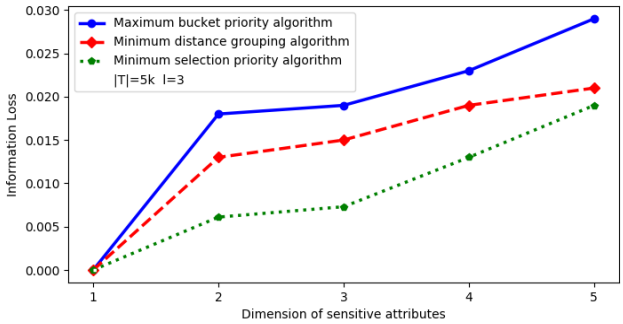


Fig. 11. Comparison of information loss under different sensitive attribute dimensions.

4.3 Execution Time Analysis

The experiments test the influence of different diversity parameter l values, t different data sets $|T|$, and different sensitive attribute dimensions d on the execution time of the three algorithms. In Fig. 12, we set the dimension of the sensitive attributes $d=2$ and $l=3$, and the execution time of the three algorithms increases with the increase of the data set. For the same data set, the execution time of the three different algorithms is gradually increase, this is because the calculation amount of the selection strategy used by the algorithm is also increased in order to obtain better grouping results, thus causing greater time cost.

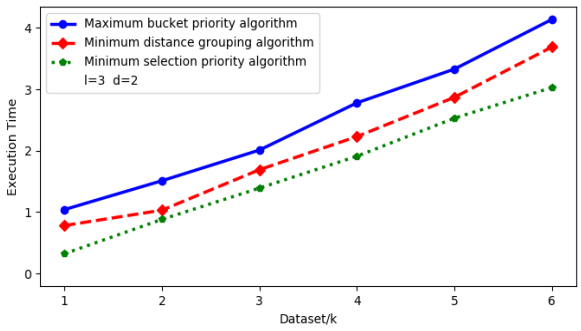


Fig. 12. Comparison of execution time under different data sets.

In Fig. 13, we set the dimension of the sensitive attributes $d=2$ and the data set $|T|=5k$. The change of l value has little effect on the execution time of the three algorithms. The analysis of the algorithms show that the minimum distance grouping algorithm's time complexity is $O(n^2)$, the minimum selection priority algorithm of the overall execution time complexity does not exceed $O(n^2)$. In Fig. 14, we set the data set $|T|=5k$ and $l=3$, the execution time of the three algorithms increase as the dimension of the sensitive attributes increase. This is because as the sensitive attribute dimensions increase, the dimension of the sensitive attribute involved in the group increases, resulting in a longer execution time of the algorithm.

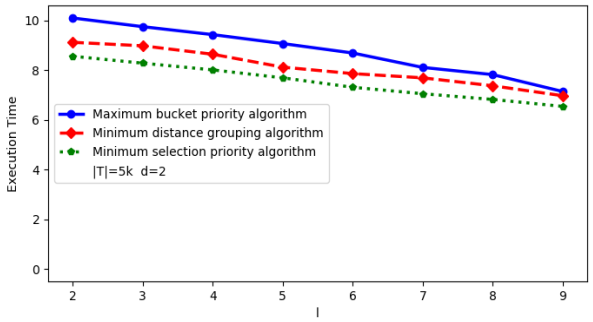


Fig. 13. Comparison of execution time under different l values.

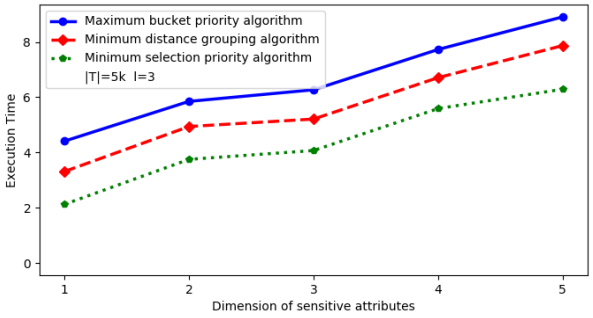


Fig. 14. Comparison of execution time under different sensitive attribute dimensions.

5. CONCLUSIONS

In this paper, the location privacy disclosure of users in mobile social network applications is studied. Based on the *l-diversity* privacy protection model, this paper proposes a location privacy protection method for multi-sensitive attributes, which protects users' location information on the client and server respectively. On the client side, the decomposition algorithm of the minimum distance grouping is adopted to conduct lightweight preprocessing for location data, so that the processed data satisfies the l_1-*diversity* principle and uploads the data to the server in forms of QIT^1 and ST^1. On the server side, the minimum selection degree priority strategy is adopted to form a personalized l_2-*diversity* group, and the processed data is uploaded in the form of QIT^2 and ST^2 (where $l_1 < l_2$), which not

only protects users' sensitive location information from being speculated by the location service provider, but also effectively avoids attackers' attacks on privacy location information. Finally, the availability and effectiveness of the two algorithms for privacy protection of location data are analyzed through experiments.

REFERENCES

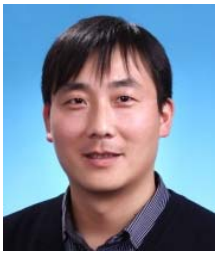
1. T. Zeng, O. Semiari, W. Saad, and M. T. Thai, "Spatial motifs for device-to-device network analysis in cellular networks," *IEEE Transactions on Communications*, Vol. 67, 2019, pp. 5474-5489.
2. Z. Huo, X. F. Meng, and Y. Huang, "Private CheckIn: Trajectory privacy-preserving for check-in services in MSNS," *Chinese Journal of Computers*, Vol. 36, 2013, pp. 716-726.
3. Y. H. Gu, J. C. Lin, and D. Guo, "Clustering-based dynamic privacy preserving method for social networks," *Journal on Communications*, Vol. 36, 2015, pp. 126-130.
4. X. Pan, X. Hao, and X. F. Meng, "Privacy preserving towards continuous query in location-based services," *Journal of Computer Research and Development*, Vol. 47, 2010, pp. 121-129.
5. H. W. Jiang, G. S. Zeng, and H. Y. Ma, "Greedy clustering-anonymity method for privacy preservation of table data publishing," *Journal of Software*, Vol. 28, 2017, pp. 341-351.
6. C. Li, L. H. Yin, K. Gen, and B. X. Fang, "Location privacy preservation approach towards to content sharing on mobile online social network," *Journal on Communications*, Vol. 37, 2016, pp. 31-41.
7. X. Liu, Q. Xie, and L. Wang, "Personalized extended (α, k)-anonymity model for privacy preserving data publishing," *Concurrency and Computation: Practice and Experience*, Vol. 29, 2017, p. e3886.
8. H. Liu, X. H. Li, B. Luo, Y. W. Wang, Y. B. Ren, J. F. Ma, and H. F. Ding, "Distributed *k-Anonymity* location privacy protection scheme based on blockchain," *Chinese Journal of Computers*, Vol. 42, 2017, pp. 942-960.
9. X. Li, S. P. Liu, F. Wu, S. Kumari, and J. J. P. C. Rodrigues, "Privacy preserving data aggregation scheme for mobile edge computing assisted IoT applications," *IEEE Internet of Things Journal*, Vol. 6, 2019, pp. 4755-4763.
10. M. Z. Cao, L. L. Zhang, X. H. Bi, and K. Zhao, "Personalized (α, l)-*diversity k*-anonymity model for privacy preservation," *Computer Science*, Vol. 45, 2018, pp. 180-186.
11. J. Liao, Z. H. Jiang, C. Guo, and Y. Ping, "Classification anonymity algorithm based on weight attributes entropy," *Computer Science*, Vol. 44, 2017, pp. 42-46.
12. T. Zhang, L. L. Zheng, Y. Z. Wang, Y. L. Shen, N. Xi, J. F. Ma, and J. M. Yong, "Trustworthy service composition with secure data transmission in sensor networks," in *World Wide Web*, Vol. 21, 2018, pp. 185-200.
13. S. B. Zhang, X. Li, Z. Y. Tan, T. Peng, and G. J Wang, "A caching and spatial *K*-anonymity driven privacy enhancement scheme in continuous location-based services," *Future Generation Computer Systems*, Vol. 94, 2019, pp. 40-50.
14. Y. Tian, X. Li, A. K. Sangaiah, E. Ngai, Z. Song, L. S. Zhang, and W. D. Wang, "Pri-

vacy-preserving scheme in social participatory sensing based on secure multi-party cooperation," *Computer Communications*, Vol. 119, 2018, pp. 167-178.
15. A. Machanavajjhala, J. Gehrke, and D. Kefer, "*l-diversity*: Privacy beyond *k*-anonymity," in *Proceedings of the 22nd International Conference on Data Engineering*, 2006, pp. 15-24.
16. X. X. Sun, M. Li, and H. Wang, "A family of enchanced *l-diversity* models for privacy preserving data publishing," *Future Generation Computer Systems*, Vol. 27, 2011, pp. 348-356.
17. A. Sadilek, H. A. Kautz, and J. P. Bigham, "Finding your friends and following them to where you are," in *Proceedings of the 5th International Conference on Web Search and Data Mining*, 2012, pp. 723-732.
18. S. Mascetti, D. Freni, C. Bettini, X. S. Wang, and S. Jajodia, "Privacy in geo-social networks: Proximity notification with untrusted service providers and curious buddies," *The VLDB Journal*, Vol. 20, 2011, pp. 541-566.
19. S. H. Zhou, F. Li, Y. F. Tao, and X. K. Xiao, "Privacy preservation in database application: A survey," *Chinese Journal of Computers*, Vol. 32, 2009, pp. 847-861.
20. X. C. Yang, Y. Z. Wang, and B. Wang, "Privacy preserving approaches for multiple sensitive attributes in data publishing," *Chinese Journal of Computers*, Vol. 31, 2008, pp. 574-587.
21. Y. Lu, "Research on privacy preserving data publishing for multi-sensitive attribute based on clustering," *Journal of Nanjing University of Posts and Telecommunications*, Vol. 20, 2016, pp. 246-254.
22. H. Zhu, D. Y. Peng, X. Q. Tang, and H. B. Liang, "A privacy preserving scheme for continuous data publishing based on clustering and segmentation technology," *Journal of Information Security Research*, Vol. 3, 2017, pp. 893-901.
23. J. M. Han, J. Yu, H. Q. Yu, and J. Jia, "A multi level F diversity model for numerical sensitive attributes," *Journal of Computer Research and Development*, Vol. 48, 2011, pp. 147-158.
24. E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," in *Proceedings of the 17th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2011, pp. 1082-1090.

Hong-Tao Li (李洪涛) received his Ph.D. degree in Computer Science from Xidian University in 2015, China. Since January 2018, he has been an Assistant Professor at the School of Mathematics and Computer Science and Technology, Shanxi Normal University. His research interests include network and information security, privacy protection, Internet of Things.

Lin-Xia Gong (巩林霞) received her M.S. degree in School of Mathematics and Computer Science and Technology, Shanxi Normal University. Her research interests include privacy protection, trust management and information security.

Feng Guo (郭锋) received his Ph.D. degree in Shanghai University in 2018. He is a Lecturer of Linyi University. His research interests include signal processing for communications, information security and privacy protection.



Quan-Li Miao (苗全利) received his M.S. degree in Shanghai Maritime University in 2011. His research interests include network and information security, privacy protection, signal processing for communications.

Jie Wang (王洁) received her Ph.D. degree in Beijing University of Technology. She is an Associate Professor in Shanxi Normal University. Her current research interests include rational secret sharing and big data privacy protection.

Tao Zhang (张涛) received his M.S. and Ph.D. degrees in Computer Science from Xidian University, China in 2011 and 2015, respectively. Since August 2015, he has been an Assistant Professor at the School of Computer Science and Technology, Xidian University. His research interests include service-oriented computing, Internet of Things, trust management and information security.