A location privacy protection method based on location semantics over road networks

Jiancheng Hu*, Jiawei Yuan
Faculty of Information Technology, Beijing University of Technology, Beijing, China

ABSTRACT

Aiming at solving the problem that it is difficult for users to resist location semantic inference attacks when they consume location-based services (LBS) in a road network environment, a location privacy protection method based on the semantic similarity of road under user collaboration is proposed: For LBS consumption, in the roads provided by collaborative users, we use the semantic similarity between each road and the road where the requesting user is located to determine the most suitable road, build an anonymous road set, and pass the agent. The method is theoretically analysed, and experiments are carried out on the anonymous success rate and anonymous time based on the Brinkhoff road network data generator. Experimental results show that this method has better privacy protection and higher service quality.

Keywords: Location privacy protection, road networks, user collaboration, semantic inference

1. INTRODUCTION

With the popularization of smart terminals and the development of positioning technology, LBS has become an indispensable part of daily life[1]. However, while users enjoy the convenience brought by LBS, their privacy will be exposed to the risk of leakage, because the data submitted by users may be collected and abused by location service providers (LSP)[2]. Therefore, how to protect users' location privacy while enjoying the convenient services brought by LBS has become an urgent problem to be solved, and it is also a hot issue that many scholars pay attention to.

There are two main structures for privacy protection methods for location services: central server structure based on a fully-trusted third party (TTP) and distributed peer to peer (P2P) structure.

The central server structure refers to the generalization, obscuration or hiding of the user's location through a trusted third-party server to achieve the protection of location privacy. It was first applied to the position k-anonymity model proposed by Gruteser et al.[3] in 2003. In this model, the applicant submits at least k locations to the server, which contains their real location, so that the attacker cannot quickly and accurately identify the applicant among the k locations, and the location is said to satisfy location k-anonymity. Since then, based on the central server structure and k-anonymity mechanism, a large number of protection methods have been proposed. However, there is no clear conclusion about the credibility of the central server. At the same time, the central server is vulnerable to attacks, and there are performance bottlenecks[4]. Therefore, more research focus has shifted to the P2P structure.

The P2P structure was first proposed by Chow et al.[5], that is, self-organizing networks, proxy users are used instead of requesting users to send query requests. In the SpaceTwice solution proposed by Yiu et al.[6] based on the P2P structure, the user sends an incremental neighbor query to the LBS server through a randomly selected anchor point, and then returns the query result to the querying user. However, because k-anonymity is not implemented, the result is the user's location privacy cannot be guaranteed. Reference[7] proposed the CoPrava-cy scheme, which combines user collaboration and incremental neighbor query. This scheme achieves the effect of k-anonymity through user collaboration, and improves the protection of user location privacy without affecting the quality of the query, but it is not suitable for collaboration between untrusted users.

Most of the above-mentioned researches are based on Euclidean space. In real life, users are faced with a complex and changeable real road network environment.

The existing location privacy protection methods in the road network environment are mainly based on K-anonymity and

*Paddington.HU@outlook.com

International Conference on Computer Application and Information Security (ICCAIS 2021),
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12260, 122601L · © The Authors.
Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2637640

road L-diversity, that is, anonymity concentration not only meets the K-anonymity needs, but also includes at least L different roads[8]. Like in references[9-11], K-L anonymity method is used to construct anonymity sets, but in the road network environment, location semantic inference attacks have brought varying degrees of privacy leakage to the anonymity sets. For example, in general, hospitals are more sensitive location information for users. If the location semantics of anonymous set are mostly concentrated near hospitals and pharmacies. Then the attacker uses semantic inference attack to locate the probability that the user is located in the hospital increases, this exposes the privacy of users to a certain extent.

Therefore, in the road network environment, the location privacy protection method is still unable to well resist semantic inference attacks, and trusted third parties may not be trusted. This article is based on the P2P structure composed of user self-cooperation. For the snapshot query in location services, a new location privacy protection scheme that supports user personalization is proposed. An anonymous road set is constructed through each collaborative user providing roads with high semantic similarity to the road where the requesting user is located. Then, let the proxy user communicates with the location service provider to reduce the risk of privacy leakage.

2. SYSTEM ARCHITECTURE

This section introduces the system structure and related definitions used by the method.

2.1. System structure

The system structure of the location privacy protection method proposed in this paper is composed of a mobile terminal and a location server, as shown in Figure 1. The mobile terminal has basic positioning and communication capabilities. The communication capability refers to the ability to communicate with the location server and also perform P2P communication with other users. Among them, the user who initiates the query request is the requesting user, denoted by U, receives the query request information from user U and submits the request to the location server is the proxy user, denoted by U_a. The location server provides services such as location-based point-of-interest query.

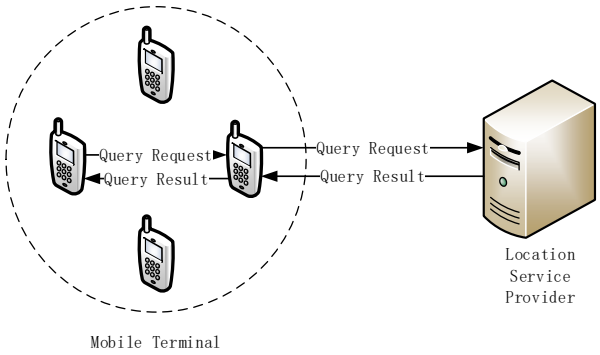


Figure 1. System structure.

2.2. Semantic road network model

The core of the method in this paper is to hide the real location of the requesting user in a set of multiple roads, and each road contains one or more semantic location information, so the road network is abstracted into an undirected graph. The vertices represent road intersections, the edges of the undirected graph represent a certain road, and there are one or more semantic locations on each edge.

Definition 1. Semantic Location. Use sl (lid, rid, x, y, st, UV) to represent the semantic location on the road segment, lid is the number of the location, rid represents the road number where the location is located, (x, y) represents the geographic coordinates of the location, st represents the semantic type of the location. All semantic locations are divided into n types, and ST = $\{st_1, st_2, \ldots, st_n\}$ is used to represent the set of n semantic location types. UV = $\{N_{st}^{t0}, N_{st}^{t1}, \ldots, N_{st}^{23}\}$ represents the number of visits by tourists of this type of semantic location in a day (divided into 24-hour periods), and the distribution of the flow of people in different time periods of each type of semantic location can be known.

Definition 2. Semantic Road Network Model. The road network is represented by an undirected graph G = (V, E), V is the set of vertices in the graph, V = $\{v_1, v_2, \ldots, v_n\}$, representing the road intersection; E is the set of edges in the graph, E =

$\{r_1, r_2, \ldots, r_n\}$, represent roads; each road r_i $(r_{id}, v_s, v_e) \in E$ is an edge in the road network, where r_{id} is the number of the road, and v_s and v_e respectively represent the start and end points of the road ($v_s \in V$, $v_e \in V$); each semantic location sl_i exists on a road. Semantic road network model as shown in Figure 2.

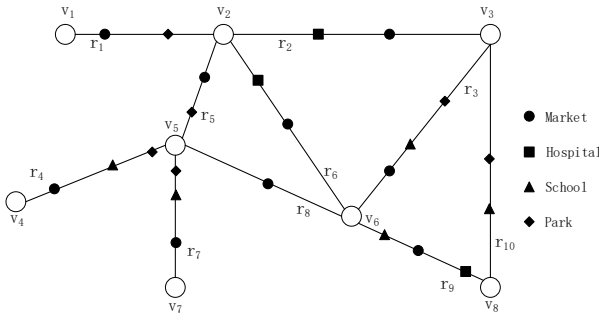


Figure 2. Semantic road network model.

3. ANONYMOUS ROAD SET GENERATION ALGORITHM

3.1. Overview to algorithm

The general steps of the method in this paper are as follows. When the requesting user U ask for snapshot query with the privacy requirement, the first step is to construct the anonymous road segment set (ARS). U first adds its own road r_u into the ARS. Then find cooperative users through broadcasting, receive the road information transmitted by the cooperative users and generate the initial cooperative user list, filter according to the semantic similarity of each road with the road r_u, add the road that meet the requirements into the ARS, and judge whether the ARS is satisfy the privacy requirement UPD. If it is satisfied, the anonymous road set is successfully constructed. If it is not satisfied, the collaborative user is required to continue to look for the collaborative user to obtain sufficient road information for screening until the demand is met or the time is up. Finally, after the ARS constructed, the user U randomly selects a proxy user U_a from the initial collaborative user list and packages the ARS and the query content into a query request Q and sends it to U_a. U_a communicates with the location service provider for service and obtains the query result, return it to user U, and U can refine the result.

Definition 3. Privacy Requirement. For the user who performs the location snapshot query service, its privacy requirement can be expressed by UPD(lnum, stnum, mwt), where lnum indicates that the final set of anonymous roads should contain at least lnum different roads; stnum indicates that the final anonymous road set should contain at least stnum different semantic location types; mwt indicates the maximum tolerance time (in seconds), if the time to construct the anonymous road section set exceeds mwt, That fails. In this method, the number of users on a road is unknown, and requesting user will eventually find a proxy user to visit the location service provider. Therefore, in this method, there is no need to specify the number of users for an anonymous road set.

Definition 4. Anonymous Road Set. The anonymous road set ARS is composed of several roads including the road r_u where the requesting user U is located, that is, ARS = $\{r_u, r_1, r_2, \ldots, r_n\}$, and the ARS meets the privacy requirement of the user U.

In the process of constructing an anonymous road set, choosing which road to join the set is an important issue. We expect to select those roads that contain the most similar type of semantic location on the road where the requesting user is located, at this time, the similarity of semantic location types on the two road sections reached the highest level. Therefore, we propose road semantic similarity to consider whether to select a road to join the anonymous road set.

Definition 5. Similarity of Semantic Location Type. The difference in the number of visits by tourists in each time period is used to indicate the similarity of two semantic location types. The smaller the difference, the more similar the two semantic location types. Assuming that the two semantic location types are st_i and st_j, the corresponding tourist visit numbers are $UV_i = \{N_i^{t0}, N_i^{t1}, \ldots, N_i^{t23}\}$ and $UV_j = \{N_j^{t0}, N_j^{t1}, \ldots, N_j^{t23}\}$, we regard UV_i and UV_j as two 24-dimensional vectors respectively, and quantify the similarity of the two types by calculating the cosine of the angle between the two vectors. The value range of similarity is [0,1], and the closer the similarity is to 1, the more similar the two types are. The calculation of similarity is shown as follows:

$$Sim(st_i, st_j) = \cos(UV_i, UV_j) = \frac{\overrightarrow{UV_i} \cdot \overrightarrow{UV_j}}{|\overrightarrow{UV_i}| \cdot |\overrightarrow{UV_j}|} \quad (1)$$

Definition 6. Road Semantic Similarity. We calculate the similarity of semantic location type of all the semantic location types in a certain road r_v with all the semantic location types of the road r_u where the requesting user is located, and then multiply the proportion of the semantic location in the road r_v by the corresponding semantic location type similarity. Finally, add up to get the road semantic similarity Rsim of the road r_v, the larger Rsim is, the higher the similarity between the road and the road r_u is. The calculation of road semantic similarity is shown as follows:

$$Rsim(r_u, r_v) = \sum_{i=1}^{|st_v|} \frac{|sl_v^i|}{|sl_v|} \cdot \sum_{j=1}^{|st_u|} Sim(st_v^i, st_u^j) \quad (2)$$

3.2. Details to algorithm

The core of the method in this paper is the construction of anonymous road set. The main thing to be solved is to select which road are added to the anonymous road set so that the constructed road anonymous set contains both the number of roads and the number of semantic location types. And the similarity of semantic location types in the road set is relatively high, to resist semantic inference attacks. Therefore, based on the comparison of the semantic similarity of roads, we propose an anonymous road set generation algorithm, as shown in Figure 3.

Input: r_u, UPD(lnum, stnum, mwt)
Output: ARS
1: Initialize the ARS, the semantic location type counter in ARS:STcount = 0, the initially collaborative user list:IUS, the firstly collaborative user list:OUS, the secondary collaborative user list:TUS, the global timer:TC = 0.
2: ARS = {r_u};
3: U send broadcast;
4: rs_p = all roads received by U from the collaborative user U_p;
5: IUS = {U_{p1}, U_{p2} , …, U_{pn}};
6: for r_i in rs_p
7: if !ARS.contains(r_i)
8: OUS.add(U_{pi});
9: map.add(r_i, Rsim(r_u, r_i));
10: end if
11: else
12: TUS.add(U_{pi});
13: end else
14: end for
15: sort(map.Rsim);
16: delete(map.Rsim < 0.5);
17: add the first (lnum-1) roads with larger into ARS;
18: STcount is updated to the number of existing semantic location types in ARS;
19: if ARS.size() < lnum && STcount < stnum
20: U informs users in TUS to send a broadcast to find cooperative users U_p';
21: rs_p' = all roads received by U from the collaborative user U_p';
22: repeat steps 6-18;
23: end if
24: if ARS.size() ≥ lnum && STcount ≥ stnum
25: return ARS;
26: end if

Figure 3. Anonymous road set generation algorithm.

The workflow of the anonymous road set generation is depicted as follows:

(1) The system receives the road r_u where the requesting user U is located and privacy requirement UPD(lnum, stnum, mwt).

(2) Add the road r_u to the anonymous road set ARS.

(3) The U sends a broadcast message to find the cooperative users within the one-hop communication range.

(4) The users U_{p1} who have received the user U's request broadcast message send the road they are located to the user U.

(5) The user U sequentially receives the road information sent by the cooperative user, and generates an initial list of

cooperative users. First discard the existing roads in ARS, copy the corresponding user information to the secondary collaborative user list, and then calculate the road semantic similarity Rsim between each remaining road and r_u, and add the different front (lnum-1) roads with larger similarity ($\geq$0.5) into ARS. And maintain its corresponding user in the first collaborative user list. At this time, check whether the ARS meets the privacy requirements according to lnum and stnum, and return to ARS directly if it meets the requirements.

(6) If it is not satisfied, notify the users in the secondary collaboration user list to broadcast to find the collaborating user again, and send the received road information to the user U, and U continues to step (5) for screening. If it is still not satisfied, it returns fail to construct ARS. In addition, in the construction process, if the time spent has exceeded mwt, it also returns fail to construct ARS.

After generating the anonymous road set ARS, the requesting user U needs to randomly select a cooperating user from the IUS as the proxy user U_a, and then U to package the ARS and the query content con into a query request Q and send it to U_a, Ua then communicate with the location service provider LSP, the LSP performs the query processing of the request content based on the ARS, generates a result set and returns it to U_a, and U_a returns the result set to U again, and finally U refines the result according to where it is.

3.3. Analysis to algorithm

3.3.1. Privacy protection analysis. The privacy protection analysis mainly starts from the attacker's point, to analyze the strength of the algorithm's privacy protection to the requesting user under the attacker's attack means. In the process of the requesting user consume the service, the attacker can intercept user information in two ways: 1. The attacker intercepts the query request and the result on the side of the location service provider; 2. The attacker pretends to be a proxy user to steal the request and infer the requesting user privacy information. The two situations are analyzed separately below:

(1) The attacker intercepts information on the LSP side. Assuming that the attacker intercepts the query request and query result at the LSP, the attacker can obtain the user's anonymous road set and query content and other information. First, the set of anonymous roads does not contain the precise location of any user. Secondly, for the requesting user, the attacker cannot directly associate this information with it without other background knowledge, so the requesting user's private information is protected; For the proxy user, the road set may contain the road information on which it is located, but the attacker cannot know the real location of the proxy user through semantic inference attacks or other means. Therefore, the privacy of the proxy user's location is protected. The query content has nothing to do with it, so there is no disclosure of the query content for the proxy user.

(2) The attacker pretends to be a proxy user. First of all, it is very difficult for an attacker to pretend to be a proxy user of a specific requesting user. When an attacker, as an ordinary collaborative user, becomes a proxy user and then executes an attack, he intercepts the proxy request packet, and the attacker can obtain the anonymous road set and query content generated by the requesting user. Also, without other background knowledge, attackers cannot know the identity information of the requesting user, nor can they know the real location of the requesting user through semantic inference attacks or other means. Therefore, the privacy of the query content and location privacy of the requesting user are protected.

3.3.2. Service quality analysis. The service quality of the location privacy protection method proposed in this paper depends on the speed of generating an anonymous road set and the communication overhead of the system. The speed of generating an anonymous road segment set depends on the speed of finding enough cooperating users. This method searches for cooperating users on the basis of finding cooperating users by itself, and rebroadcasts them to find cooperating users again. Therefore, it takes a short time to find a set of roads that can meet the requirements of the requesting user. In addition, the communication overhead of the system consists of the communication between the requesting user and the cooperating user, the communication between the requesting user and the proxy user, and the communication between the proxy user and the LSP. Since some collaborative users are obtained by the collaborative user broadcast, it can be considered that part of the communication overhead of the requesting user is allocated to some collaborative users. The communication overhead of the requesting user with the LSP is also transferred to the proxy user. It only needs to maintain a connection with users within the communication range of one hop to receive information, send query packets, and receive result packets, so the communication overhead is relatively small.

4. EXPERIMENT

The experimental hardware platform of this article is CPU 2.5GHz, 16GB memory, and Microsoft Windows 10 operating system. The algorithm is implemented in Java.

4.1. Experimental data set and parameter settings

The road network data adopts the road network data of Oldenburg, Germany. The road network contains 7,035 road intersections and 6,105 roads, and the road network data generator of Brinkhoff[12] is used to generate moving object data. The experimental default parameters are shown in Table 1.

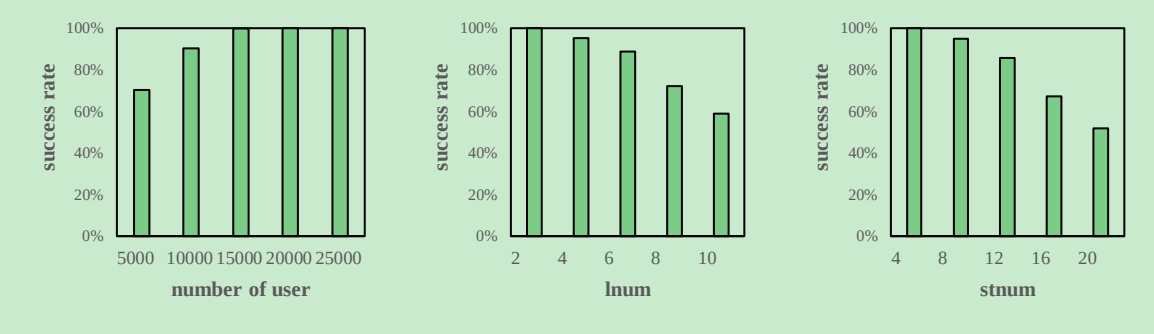
Table 1. Parameter settings.

Parameter	Default	Range
Number of users	10000	5000-25000
lnum	5	3-10
stnum	10	5-20
mwt	3	
Number of semantic location type	20	
Number of points of interest	1000	
Number of the requesting users	1000	

4.2. Analysis of results

This paper mainly verifies the effectiveness of the algorithm from two aspects: the success rate of generating anonymous road set and the average time spent.

4.2.1. Set generation success Rate. Figure 4 shows the change in the success rate of set generation relative to the total number of users on the road network, the demand number of road, and the demand number of semantic type.



(a) success rate and user number (b) success rate and lnum (c) success rate and stnum

Figure 4. ARS generation success rate.

As Figure 4a shows, the more users on the road network, the higher the success rate of set generation. It is easy to understand, because the more the number of total users, the relative increase in the number of users on each road, the faster the search for collaborative users, the faster the collection will be generated. Conversely, when the number of total users is small to a certain extent, then the number of users on each road is relatively reduced, and it takes more time to generate the collection, and it may even fail to find enough collaborative users, causing the generation to fail. In Figures 4b-4c, with the stricter privacy requirement, the success rate of set generation decreases. When the number of total users of the road network is set to 10 000 by default, the generation collection time is set to 3 seconds by default, and the demand number of road and the demand number of semantic type are 8 and 16, respectively, the success rate drops significantly. Therefore, when the density of road network users is low, privacy requirements should not be set too strict.

4.2.2. Average generation time. According to Figure 5, when the number of total users on the road network is constant, as the number of road and semantic type with privacy requirement increases, the average generation execution time will increase because more collaborative users need to be found. Under the same conditions, the more the total number of total users on the road network, the shorter the average generation execution time. The reason is also: the more the number of total users, the relative increase in the number of users on each road, the faster the search for collaborative users, the faster the collection will be generated.

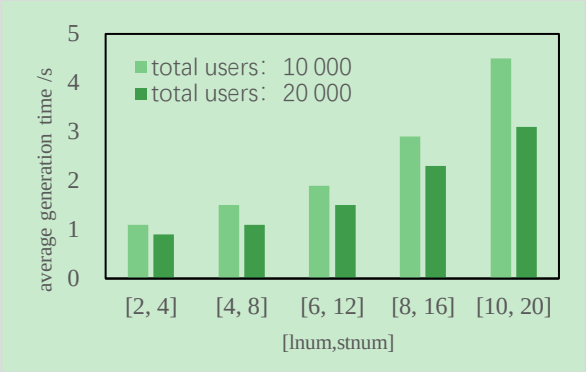


Figure 5. ARS generation average execution time.

5. CONCLUSION

This paper presents a road network location privacy protection method based on user collaboration, which achieves a certain balance between personalized location privacy protection and service quality when users on real road networks perform LBS consumption services. Due to factors such as untrustworthiness and performance bottlenecks in the central server and easy targets for attacks, this paper adopts a P2P structure. Based on the collaboration of users in self-organized networks, the roads of each user are obtained, and anonymous road set are generated based on the semantic similarity of roads. At the same time, the proxy user is selected to communicate with the LSP to protect the privacy of the requesting user, and finally obtain the query result to use. Through a large number of experiments based on real data and simulated data, the feasibility and effectiveness of the method in this paper are proved.

REFERENCES

[1] Wang, B., Yang, X., Wang, G., Yu, G., Zang, W. and Yu, M., "Energy efficient approximate self-adaptive data collection in wireless sensor networks," Frontiers of Computer Science, 10, 936 (2016).
[2] Wang, L. and Meng, X. F., "Location privacy preservation in big data era: A survey," Journal of Software, 25, 693 (2014).
[3] Gruteser, M. and Grunwald, D., "Onymous usage of location-based services through spatial and temporal cloaking," Proc. of the 1st Inter. Conf. on Mobile systems, Applications and Services, 31 (2003).
[4] Mao, D. H., Cai, Q. and Li, H., "LBS privacy protection based on user collaboration under road network conditions," High-tech Communication, 23, 1148 (2013).
[5] Chow, C. Y., Mokbel, M. F. and Liu, X., "A peer-to-peer spatial cloaking algorithm for anonymous location-based service," Proc. of the 14th Annual ACM Inter. Symp. on Advances in Geographic Information Systems, 171 (2006).
[6] Yiu, M. L., Jensen, C. S. and Huang, X. G., "SpaceTwist: Managing the trade-offs among location privacy, query performance, and query accuracy in mobile services," Proc. of the 24th Inter. Conf. on Data Engineering, 366 (2008).
[7] Huang, Y., Huo, Z. and Meng, X. F., "CoPrivacy: A collaborative location privacy-preserving method without cloaking region," Chinese Journal of Computers, 34, 1976 (2011).
[8] Li, M. and Qin, Z. G., "Research progress of location privacy protection technology in road network environment," Computer Application Research, 31, 2576 (2014).
[9] Wang, Y., Xia, Y. and Hou, J., "A fast privacy-preserving framework for continuous location-based queries in

road networks," Journal of Network & Computer Applications, 53, 57 (2015).
[10] Chow, C. Y., Mokbel, M. F. and Bao, J., "Query-aware location anonymization for road-networks," GeoInformatica, 15, 571 (2011).
[11] Xu, E. J. and Liu, X. Y., "A location privacy preserving approach on road network," Chinese Journal of Compute, 34, 865 (2011).
[12] Brinkhoff, T., "A framework for generating network-based moving objects," Geoinformatica, 6, 153 (2002).