A LOCATION PRIVACY PRESERVING ALGORITHM FOR MOBILE LBS

Wei Li, Wei Jiao, Guangye Li

School of Computer Science and Engineering,
Beihang University, Key Labratory of Beijing Network Technology, Beijing 100191, China
Chinaliw@buaa.edu.cn, plakjld@gmail.com, leeguangye@163.com

Abstract: Location-Based Service (LBS) combined with mobile devices and Internet become more and more popular, and are widely used in traffic navigation, intelligent logistics and the point of interest query. However, most users worry about their privacy when using the LBS because they should provide their accurate location and query content to the untrustworthy server. Therefore, how to protect users' privacy is very important to the mobile LBS application. Continuous queries tracking attack is one of the typical attack models in mobile LBS. This paper analyzes the query association attack model for the continuous query in mobile LBS, formalizes the background knowledge of attackers considering the spatiotemporal dimension. In order to resist the query association attack for continuous query in mobile LBS, this paper improves the k-sharing model and proposes the location distribution aware cloaking algorithm which is distribution-aware when generating cloak regions, it satisfies the m-invariant and l-diversity. Finally, a set of experiments show the effectiveness of the approach.

Keywords: Location-based service; Privacy protection; Continuous query

1 Introduction

With the development of the ubiquitous wireless technology and mobile positioning technologies, there is explosive growth of location-based services (LBS) in recent years [1-2]. Mobile clients can issue queries together with their accurate location information and query contents to request LBS. However, the privacy of user's location and query content information may be threatened by the untrustworthy servers. The disclosure of user's location information and query contents are possible to lead to the disclosure of users' behavior patterns, health status, physical stalking, and personal privacy information [1-3].

Continuous queries tracking attack is one of the most typical attack models [4-9]. Most query tracking models usually consider the background knowledge of the attack on temporal dimension, which means that the attacker has users' information at different moments of cloaking regions. Moreover, the attacker can also have anonymity related information of different users on the spatial dimension. Therefore, the attacker could effectively infer the specified users related to the location when combining the spatiotemporal information.

This paper analyzed the characteristics of the query attacks for continuous queries in mobile LBS, formalized the attacker's background knowledge in both horizontal spatial dimension and vertical temporal dimension. In order to resist the query association attack for continuous query in mobile LBS, it improves the k-sharing model and proposes the location distribution aware cloaking algorithm which is distribution-aware when generating cloak regions, it satisfies the m-invariant [9-10] and l-diversity [9, 11-12]. Therefore, the attacker cannot distinguish the position and content of the user in continuous query.

The rest of the paper is organized as follows: Section 2 analyzes the related research work on the privacy protection for continuous query. Section 3 presents the system model and attack model; Section 4 presents location distribution aware cloaking algorithm; Section 5 verifies the effectiveness of the algorithm through experiments; Finally, Section 6 presents the conclusion and future research.

2 Related work

In this section, we review the relevant work on location privacy protection of continuous queries in LBS. Chow [2] first proposed privacy protection method for the continuous queries. This method focused on the two forms of attacks: Query sampling attacks and query tracking attacks. In order to resist these two kinds of the attack models, the paper presented the cloaking algorithm which can make the cloaking region meet the k-sharing property. So the cloaking region can effectively prevent query sampling attacks and query tracking attacks. However, due to the mobility of the users, anonymous area generated from the cloaking algorithm tends to be too large to guarantee the quality of service.

Toby Xu [13] presented that the frequently updated users' location information in the continuous queries may result in the leakage of the user's location privacy. This paper evaluated the protection degree with entropy, and proposed two algorithms of plainKAA, and advanceKAA. The anonymous area produced by PlainKAA algorithm will be too large or too small. In

978-1-4673-1857-0/12/$31.00 ©2012 IEEE

order to reduce the cloaking region, the algorithm of advanceKAA generates anonymous area based on the sets of users containing in the last anonymous area, from the last anonymous area to calculate the candidate collection, until the smallest cloaking region satisfied privacy requirements can be found.

Pan [8] studied both current location and future location of mobile users, and proposed the δ_p privacy model and δ_q quality model for the proximity of the location changing caused by the user's movement. In order to balance the tradeoff between privacy protection and service quality, a greedy anonymity algorithm (GCA) was proposed utilizing the similarity of two objects, to protect the user's location privacy and query privacy.

Rinku Dewri [9] believed that k-anonymity model and l-diversity model failed to provide an effective response for attack to the query. Query l-diversity guarantees at least *l* distinct values in every query, but does not try to always keep the same set of values in the queries. Therefore, they analyzed the risk of loss of privacy in continuous query, and introduced the m-invariant model to protect data privacy, including the query m-invariant model and m-invariant Cloak algorithm. However, this method needed to send a large number of repeated queries, which may lead to the poor quality of service.

Our solution considers both the user location distribution and query distribution, and presents a location distribution aware cloaking algorithm to generate the cloaking region. The cloaking region cannot only satisfy the m-invariant and l-diversity, but also satisfy the k-sharing property. So the attacker cannot inference specific user location and query contents for continuous query in LBS. Meanwhile, this method can achieve a balance between privacy protection and privacy security.

3 Attack model

The trusted third party architecture [2, 6-7, 14] is used in our system, and the centralized anonymity server is responsible for cloaking the locations of mobile users. In this section, a query association attack model is formalized. The type and contents of user's query may change over time in LBS. Considering the characteristics of attacks associated with the continuous queries in LBS; the anonymous space must meet the k-sharing feature, which is the basis of the user location distribution aware cloaking algorithm.

3.1 Background of the attacker's knowledge

Success or failure for attacker largely depends on the background knowledge they have. This paper assumes that the attacker can obtain location information of all users during the users' query in cloaked area. Therefore, the background knowledge of the attacker can be formalized as follows.

Definition 1: (Background Knowledge, BK) $P = \{p_i \mid 1 \le i \le duration(u)\}$, where $duration\ (u)$ is the duration of continuous queries issued by the user *u*. p_i consists of three parts logically: (*id*, *loc*, *t*). The *id* represents the user's identity. The *loc* is the user's specific location. The *t* is the time stamp. The attacker has background knowledge of the location information of a user at time *t*.

3.2 Query association attack model

There is an example under the following condition which is shown as Figure 1.

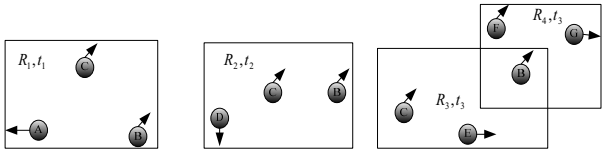

Figure 1 Cloaked region of user *C* at different time

Figure 1 illustrates the cloaked region *R* of user C at different time t_i. The black spots in this figure represent a user. Cloaking region satisfies the privacy model of m-invariant and l-diversity. At time t_1, the users set corresponding R_1 is {A, B, C}, and the query content set is {a, b, c}, the moving direction of A is significantly different with B and C. At time t_2, the users set corresponding R_2 is {B, C, D}, and the query content set is {b, c, d}. At the same time, the moving direction of D is significantly different with B and C. At t_3, the users set corresponding to R_3 is {b, c, e}. The users set corresponding R_4 is {B, G, F}, and the query content set is {b, g, f}.

Due to the different patterns of the users' movement, the user set of a user's cloaking region would continuously changes. Attacker could utilize such feature to infer the user's query content [6]. The combination of m-invariant and l-diversity models could not effectively cope with such attacks. In the above example, the attacker is able to accurately infer the query of user B and C through using the characteristics of continuous queries, that b as the query content of B and c as the query content C because of the overlap region between R3 and R4.

Definition 2: (Session Information, SI) Assume the anonymous space of user u at $t_1, t_2, \ldots t_n$ is $CR_1, \ldots, CR_n$, server sets is $S_1, S_2, \ldots, S_n$. $SI = \bigcup_{i=1}^{i=n} \{t_i\} \times \{CR_i\} \times S_i$.

Definition 3: (Query m-Invariant) Query m-invariant means that the invariant set has a size of at least m in every query, but does not try to always keep the same set of values in the queries [9].

Definition 4: (Query Association Attack) For given SI and BK, the query association attack of user u is a mapping $f : BK(u) \rightarrow SI(u)$. In Ref. [9], there are two mapping satisfying the following conditions:

1) For arbitrary $b \in BK(u)$, we can obtain the only element $e \in SI(u)$.

2) If $b \in BK(u)$ and $f(b) = e$, it should meet

a) $b.t = e.t$;

b) $(b.loc.x, b.loc.y)$ is inside $e.CR$;

c) For all $b' \in \{b'' \in BK(u) \mid b''.u = b.u\}$, $f(b').s = e.s$;

Based on these two conditions, the mapping should also satisfy the other condition:

3) For any $e' \in SI(u)$, if the overlapping region of e and e' is C, the u query $s \in \cap SS_i$ and $SS_i = Services(e.CR, t_i) \cap Services(e'.CR, t_i)$, $u \in Users(e.CR, t_i) \cap Users(e.CR, t_i)$.

Where Services *(CR, t)* represents the corresponding to query set in the cloaking region *CR* at *t*, *Users (CR, t)* is the user set corresponding to cloaking region *CR* at *t*.

The algorithm of query association attack is as follows:

//Query Association Attack

Input：L：users' location information log

Output：users' query content

1 queryCollection = findLog(L);// look for cloaking regions from the anonymity server

2 query= doIntersection(queryCollection);//do intersection operation for each query set

3 P= calculateProbability(query);//calculate the probability corresponding to the queries

4 service= getMaxProbablity(P);//find the largest probability and return the corresponding query

5 return service;

Attacker would firstly search all cloaking regions containing users in the location server's log based on the users' location information, and then take the intersection of the query contents corresponding to the cloaking region; calculate the probability of each query. The maximum probability correspond is to the users' query.

4 Algorithm design and analysis

The query association attack is caused by the skewed distribution of users' location and query. However, when the attacker exploits the query association attack, the m-invariant privacy protection model cannot guarantee the privacy of the user's query. In order to cope with the query association attack effectively, we need to strengthen the cloaking region characteristics.

4.1 Algorithm goal

In order to cope with query association attack effectively, the shared cloaking region should have k-sharing characteristics. Not only the anonymous space of the user query set meets the m-invariant model, but also the anonymous space is allowed to be shared by its users. That is, it satisfies the characteristics of k-sharing.

Definition 5: (K-sharing) K-sharing means that a cloaked spatial region not only contains at least k users, and the region is also shared by at least k of these users.

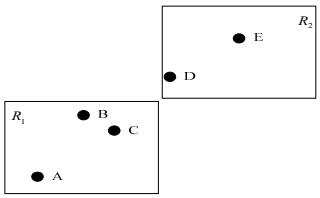

Figure 2 The anonymous areas which satisfy k-sharing

Figure 2 shows that R1 is an anonymous area which is shared by user A, B and C. R2 is an anonymous area which is shared by user D and E. Because the two anonymous areas do not have the overlapping area, the attacker cannot infer users' queries.

Theorem: If the cloaking region R satisfies the above characteristics, assume the maximum value of *k* user setting is K, then the attackers adopt query association attack which can infer the probability of a particular user query is not greater than $1/K$.

Proof: Assume the anonymous space of user *u* at $t_1, t_2, ...t_n$ is $R_1, ..., R_n$, server sets is $S_1, S_2, ..., S_n$, the user sets are $U_1, U_2 ..... U_n$.

Set $n = \left|\prod_i S_i\right|$, $p = \left|\prod_i U\right|$, thus $n \geq K$.

The service request is sent by user *u* is *s*, $s \in \prod_i S_i$, and the probability of attacker inferring the service request *s* is sent by user *u* satisfying $\frac{n^{p-1}}{n^p} = \frac{1}{n}$; because $n \geq K$, the attacker can infer the probability of a particular user query is not greater than $1/K$.

4.2 Algorithm design

// Location Distribution Aware Cloaking Algorithm

Input：t：time value the cloaking region needs to satisfy; R：user's query request set

Output：Anonymous space collection

1 L=HilbertOrder(R);

2 for(all $l \in L$){

3 $P = P \cup \{l\}$; $S = P \cup \{l.s\}$;

4 IS= IS $\cup$ getInvariantSet(l); m=getMaxM();

5 if($|IS \cap S| \geq m$ ||(($m \leq p.size$)&&(IS==null))){

6 peerGroups = split(P);

7 if(peerGroups!=null){

8 peerGroups.serviceSet = S;

9 annoResult.add(peerGroups);

10 for(all $p \in P$){

11 if(Is==null) p.invSet=S;

12 else p.invSet= $IS \cap S$;

13 }//end if

14 remove P from l; P=null; S=null; IS=null;

15 }//end if

16 }//end if

17 }//end for

18 return annoResult;

The input of the Location distribution aware cloaking

algorithm is the user request containing the user's current location, user id and user service request parameters. Function HilbertOrder() inputs the current user's request collection, and changes the request of two-dimensional coordinates into one dimensional indexing using Hilbert Curve filling algorithm. Then sorting user queries collection is based on the Hilbert index code from large to small. All the users' requests are traversed, and record the traversed users set as P. The current users' request collection services set as S. getInvariantSet () was used to obtain the same services collection IS of a specific user. If the cardinality of the intersection of IS and S greater or equal to m or IS is empty, and the number of query is greater than or equal to m, then the split method is called to generate cloaking space. Split algorithm is similar to HilbertCloak, and the smallest rectangle anonymous space MBR is not larger than the upper bound of the anonymous space.

4.3 Algorithm analysis

The algorithm needs to calculate Hilbert index value of different users for the first implementation; however the Hilbert space can be used for many times once the first calculation is completed. Therefore, the calculation time of all users Hilbert index is $O(n)$. The time complexity used to divide anonymous requests of users is $T(n)=\sum_{i=1}^{i=m} O(k_i)$, where $\sum_{i=1}^{i=m} k_i = n$, $T(n)=O(n^2)$. Therefore, time complexity of the location distribution aware cloaking algorithm is $O(n^2)$.

5 Experimental evaluation

The user trace data used in the experiments is generated by the Brinkhoff data generator [15], and the road map used the city of Oldenburg, Germany. Trajectory of 20000 users and 30000 queries are generated in the experiments, and the query contents uniformly distributed in the selected 200 data. The map space is divided into $2^{10} \times 2^{10}$ grid in the experiment, and Hilbert index value for each user is generated in the grid. We use cloaking time and query success rate as the evaluation metrics [14].

Figure 3 and Figure 4 show the comparison between m-invariant Cloak and Distribution Aware Cloaking. From Figure 3, the m-invariant Cloak has a longer cloaking time with the increasing of *m*, while Distribution Aware Cloak can achieve less cloaking time with the increasing of *m*. This is because the m-invariant Cloak has to create an anonymous space for every user, and in our Distribution Aware Cloaking *k* users share the anonymous space, thus it cost less time.

It can be seen from Figure 4, the m-invariant Cloak and Distribution Aware Cloaking have a smaller query success rate with the increasing of *m*, but Distribution Aware Cloaking's query success rate is larger than m-invariant Cloak. This is because in our Distribution Aware Cloaking, *k* users must meet the k-sharing property, but the m-invariant Cloak does not need to satisfy.

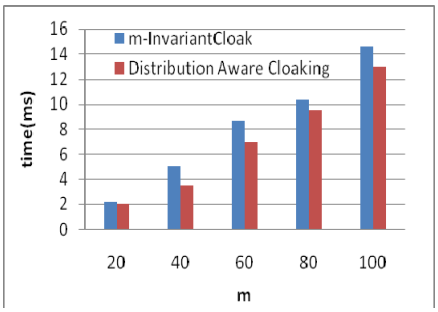

Figure 3 Comparison between m-invariant cloak and distribution aware cloaking about cloaking time

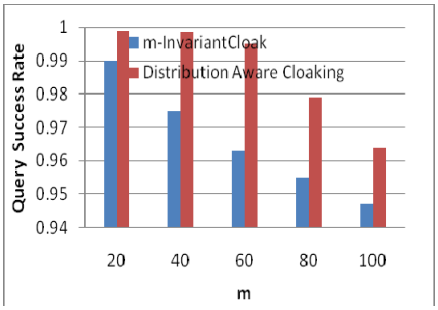

Figure 4 Comparison between m-invariant cloak and distribution aware cloaking about query success rate

Figure 5 and Figure 6 show the comparison of the algorithm performance under the condition of the anonymous space when *a* value lies in 10000m^2 and 40000m^2 respectively. Figure 5 is comparison of average relative anonymity of the area: generally speaking, the average relative anonymity area when *a*=40000 is larger than *a*=10000, because *a* value specifies the maximum area of anonymous space: the greater the upper bound anonymous space, the corresponding anonymous space is relatively large. Figure 6 is an anonymity server processing time used by each anonymous request. It can be seen that anonymity increases with the increasing m value. Therefore, with the difference of the algorithm in the anonymous time, average relative anonymous area are not obvious when *a*=10000 or *a*=40000, and the algorithm is more stable at the same time.

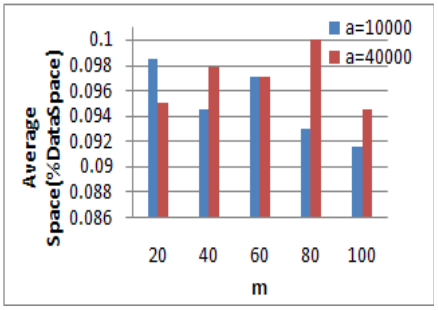

Figure 5 Comparison of average value of relative anonymous area with different *a* value

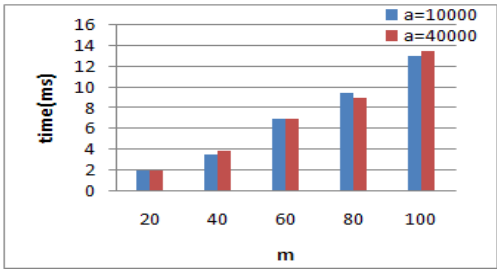

Figure 6 Comparison of anonymous time with different *a* value

Figure 7 and Figure 8 reflect the algorithm performance comparison of cloaking algorithm under different conditions for various users. The parameter *m* is fixed at 20 and the anonymous space is set to 10000m^2, and then we compare different user number. Figure 7 is comparison of the average anonymous area, and it can be seen that the user distribution tends to be more intensive with the increasing number of users, and therefore the smaller average anonymous space is generated. Figure 8 shows the comparison result of anonymous time required, which also increases with the increasing number of users.

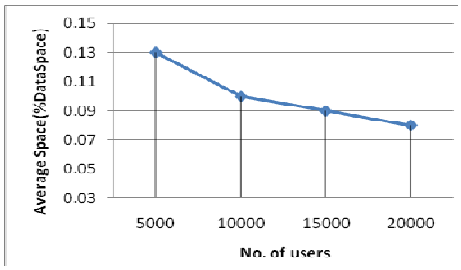

Figure 7 Comparison of average anonymous area with different user number

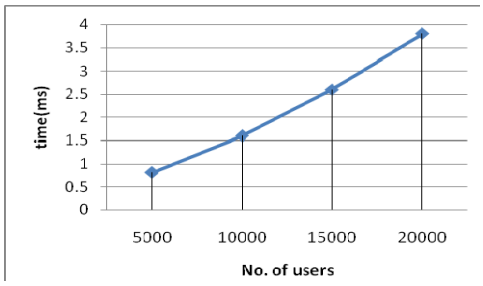

Figure 8 Comparison of average anonymous time with different user number

6 Conclusions

This paper analyzed the characteristics of query association attack model for continuous query in location-based services from the spatiotemporal dimension. In the temporal dimension, the relevance of anonymous space generated by a user in the valid query period is studied, and in the spatial dimension, the relevance of different anonymous spaces generated at the same period is compared. Then this paper formalized the query association attack model. In order to prevent query association attack effectively, this paper proposed location distribution aware cloaking algorithm, and finally verified the effectiveness and validity of cloaking algorithm by experiments. In future work, we will extend the location distribution aware cloaking algorithm from Euclidean space to road network, which can provide the effective protection of location privacy for mobile LBS application.

References

[1] Mokbel M F, Chow C, Aref W G. The new Casper: Query Processing for Location Services without Compromising Privacy. Proceedings of the International Conference on Very Large Data Bases (VLDB'06), Seoul, Korea, 2006: 763-774.

[2] Chow C, Mokbel M F. Enabling Private Continuous Queries for Revealed User Locations [C]. International Symposium on Spatial and Temporal Databases, 2007: 258-275.

[3] FoxsNews [EB/OL]. http://www.foxnews.com/story/0,2933,131487,00.html.

[4] Aoying Zhou, Bin Yang, Cheqing Jin, Qiang Ma. Location-Based Services: Architecture and Progress [J]. Chinese Journal of Computers, 2011,(07):34-46.

[5] Pingley A, Zhang N, Fu X, A H Choi, Subramaniam S, and Zhao W. Protection of query privacy for continuous location based services[C]. International Conference on Computer Communications (INFOCOM'11), 2011:1710-1718.

[6] Chow C , Mokbel M F and Liu X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer environments [J]. GeoInformatica, 2011,(15):351-380.

[7] Demiryurek U., Kashani F. B., Shahabi C. Efficient Continuous Nearest Neighbor Query in Spatial Networks Using Euclidean Restriction. International Symposium on Spatial and Temporal Databases (SSTD'09), 2009: 25-43.

[8] Xiao Pan, Xing Hao, XiaofengMeng. Privacy Preserving Towards Continuous Query in Location-Based Services [J]. Journal of Computer Research and Development, 2010:66-76.

[9] Dewri R, Ray I, Whitley D. Query m-invariance: Preventing query disclosures in continuous location-based services[C]. Proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware, 2010: 78-88.

[10] Xiao X, Tao Y. m-Invariance: Towards Privacy Preserving Republication of Dynamic Datasets[C]. Proceedings of the 2007 ACM SIGMOD International Conference on Management of Data, 2007:689–70011.

[11] Macha A, Gehrke J, Kifer D. l-diversity:Privacy beyond k-anonymity[C].Proceedings of the Intemational conference on Data Engineering, 2006: 5-14.

[12] Liu F, Hua A K, Cai Y. Query l-Diversity in Location-Based Services[C].Proceedings of the 10th International Conference on Mobile Data Management: Systems, Services and Middleware, 2009:436–442.

[13] Toby Xu, Cai T. Location anonymity in continuous location-based services[C]. Proceedings of International Symposium on Advances in Geographic Information Systems (GIS), 2007:80–90.

[14] Wei L, Guangye L, Chunlei L. Query-Aware Anonymization in Location-based Service[C]. The 7th International Conference on Computational Intelligence and Security(CIS2011), 2011:741-745.

[15] Brinkhoff T. Framework for generating networkbased moving objects [J]. GeoInformatica, 2002,6(2):153–180.