ATLANTIS
PRESS

Advances in Computer Science Research, volume 44
3rd International Conference on Wireless Communication and Sensor Network (WCSN 2016)

Enhanced k-anonymity Privacy Protection Scheme in Continuous LBS Queries

Wei Wang
Science and Technology on Communication Information
Security Control Laboratory
No.36 Research Institute of CETC, Jiaxing, China, 314033
Email: wwzwh@163.com

Wen-Hong Zhao
Nanhu College, Jiaxing University, Jiaxing, China, 314001
Email: wwlofty@gmial.com

Zhong Cheng, Xing-Hua Li
School of Cyber Engineering, Xidian University, Xi'an China, 70771

***Abstract*-The existing anonymous region construction schemes based on continuous Location-based service (LBS) queries do not solve the temporal and spatial relationship of common user sets in adjacent anonymous regions. The attacker can reduce the anonymous region and reduce the common user set by the Maximum movement boundary attack, and reduce the privacy protection level of the user. Aiming at the above problems, considering the time accessibility of adjacent anonymous region and common user set, this paper proposes a location privacy enhancement scheme based on historical footprint for k-anonymity. The security analysis shows that the scheme does not increase the computational cost and communication cost of the system while guaranteeing the privacy queries of the users, and does not affect the user's quality of service. It has good validity and practicability.**

Keywords-continuous LBS queries; reachability; privacy query; quality of service

I. INTRODUCTION

LBS is a kind of value-added service that provides the corresponding service for users according to the position submitted by the user and supported by the location service provider. Since the user is required to submit location information, the user's location information may also be leaked while enjoying the convenience brought by LBS. Using this information, an attacker can infer private information such as home address, religious belief, political inclination, etc., even threaten user privacy. Therefore, the protection of user location privacy in LBS is of great concern to researchers. Currently, the most widely used method is location k-anonymity. This technology was first proposed by Sweeney[1] in privacy protection issues with data disclosure, and then it was introduced to the user location privacy protection in LBS by Gruteser[2] and others. The basic idea is to generate an anonymous region containing at least k users instead of the true location of the originating querying user. Since the area contains at least k users, an attacker is not able to associate a query with a user with a probability of more than 1 / k.

LBS queries are typically grouped into two categories at the frequency of the user's queries. One for the snapshot query, that is, a user sends only a single LBS query at a certain time, the other type is continuous queries, that is, the user sends several queries consecutively in a certain period of time to obtain the same or different services. In continuous LBS queries, there are some temporal and spatial relationships between adjacent queries. At the same time, the same temporal and spatial relationship exists in the common user set in the adjacent anonymous region. An attacker can narrow the range of anonymous region by using the Maximum movement boundary attack, then reduce the common user set, and reduce the privacy protection level of the user, even infer the true location of the user.

The main contributions of this paper are as follows:

- Combined with the current anonymous region construction program, through analysis, with the introduction of the Maximum movement boundary attack, it is pointed out that ignoring the temporal and spatial relationship between the anonymous regions will lead to narrowing the anonymous region, thus reducing the common user set and reducing the user privacy protection level.
- Using temporal and spatial information based on historical footprint to judge the time accessibility of adjacent anonymous regions and common user sets, and proposing a privacy enhancement scheme for regional reachability.
- Verified by experiment Experiments that the common user set constructed by the existing schemes are not able to resist the Maximum movement boundary attack, resulting in the user privacy query ungratified. The scheme of this paper can fully meet the user's privacy query without increasing the computational cost and communication cost of the system and without affecting the service quality of users.

The remainder of this paper is organized as follows. First, Section 2 reviews the location privacy protection scheme in existing continuous LBS queries. In section 3, we introduce the problem of Maximum movement boundary attack in continuous queries, and propose a new continuous LBS queries privacy enhancement scheme based on region reachability. The effectiveness and practicability of the proposed scheme are given in Section 4. Finally, Section 5 summarizes the work of this paper.

II. RELATED WORK

This section briefly reviews the existing location privacy k-anonymity protection scheme for continuous LBS queries.

Aiming at location privacy threats under continuous LBS queries, Chtableow and Mokbel[3] and others first proposed and implemented an anonymous region using space anonymity to solve the problem of sampling attacks and intersection attacks under continuous queries. However, since all the queries need to use the common user set which formed the anonymous region for the first time, as the number of queries continues to grow, the anonymous region generated by the program will grow rapidly. In order to solve the problem of too large anonymous region, Xu[4] and others first proposed to solve the problem of continuous queries based on historical footprint information. They filtered out the k-1 historical footprints track record which is the closest to user and constructed the anonymous region. However, this scheme does not consider the direction of user movement. The computation is huge and the efficiency is very low. At the same time, Xu[5] and others adopted the development of public areas to represent the user's location privacy level, and proposed a privacy protection method based on sense. However, neither of these methods can meet the user-defined privacy query. However, the L2P2 scheme which proposed by Wang[6] and others, start to from the initial location of continuous queries, and expand the anonymous region for each location progressively until all the queried privacy queries are met. Li[5] and others point out that when anonymous regions are constructed using the historical footprints of other users, there is a problem that the anonymous region is too large and the quality of service is degraded. They reduce the anonymous region by suppressing a small number of LBS queries from users. Liu[7] and others for the first time use pseudo location in continuous LBS queries, where the temporal and spatial relationships of adjacent anonymous regions are also considered. But this method needs to call the map interface to judge the reachability, which greatly increases the query latency and greatly increases the communication cost.

In summary, the existing k-anonymous-based programs are not good solutions to the problem of temporal and spatial relationship between adjacent anonymous regions, resulting in the privacy accessibility greatly increase the query latency and communication cost, and seriously reduce the user's service quality.

III. PROBLEMS OF CONTINUOUS QUERIES ANONYMOUS REGION AND PRIVACY ENHANCEMENT SCHEME

Be same with the current common k-anonymity architecture, this paper is also based on the center anonymous server architecture to start discussion. Referring to the attacker model in the existing literature, this paper assumes that the attacker has the following capabilities:

- The attacker knows that the time and location information of all LBS queries submitted by users. That is, any anonymous region and user set sent from the anonymous server to the location server can be intercepted.
- The attacker with maps and background knowledge, such as the arrival time of any two points, the location of the users' distribution and so on. Considering that most schemes use historical footprint information to construct a common user set to solve the problem of intersection-seeking attack, this paper on this basis for further study.

A. *Maximum Movement Boundary Attack*

Maximum movement boundary[9] attack means that the attacker can know the distance between any two points and the maximum speed of the road because of the background knowledge of the map, so that the reachable time range between the two points can be deduced. We introduce the Maximum movement boundary attack into the existing k-anonymity scheme. Now we regardless of the historic user's temporal and spatial relationships when we adopting historic user footprint construct the anonymous region, so that an attacker can use the time interval of user's query to determine the user's range of activity. And then the attacker excludes some areas in the anonymous region, narrowing the anonymous region. If the attacker has a better understanding of the distribution of historical footprints, then the historical users of the excluded precincts will be eliminated altogether, the anonymous common user set will be narrowed, which allowing the attacker to discover the querying user at a probability of less than 1 / k.

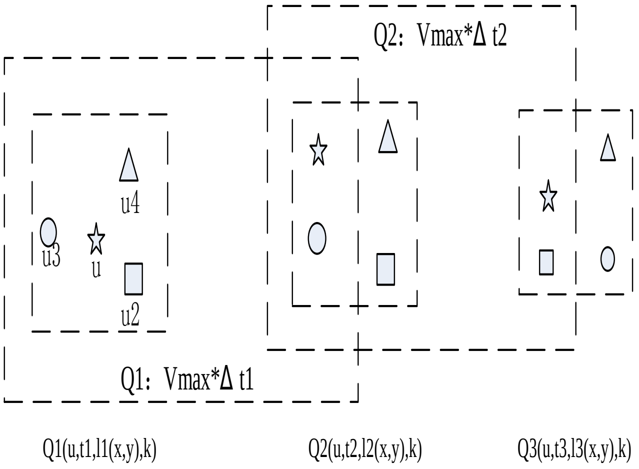

Figure 1. Maximum movement boundary attack mode

The middle of Fig. 1 is an anonymous region constructed by three consecutive LBS queries. We assume that the time interval between Q_1 and Q_2 is Δt_1, the maximum speed for the road is V_{max}. Then the part of the dashed area indicates the reachable range of the current anonymous region in the direction of the next anonymous region. That is, in the anonymous region of query Q_1 in the figure, the range of any user's activities is only a dashed box in the time interval Δt_1. Then it can be seen from the figure, in the query Q_2, the right half of the region is not able to be reached by any user in query Q_1. The attacker can determine that the partial area does not contain the querying user. Thereby excluding the partial region from the anonymous region and narrowing the anonymous region. When the situation is serious, anonymous area can be reduced to the anonymous area below the minimum A_{min} Obviously, it does not meet the privacy needs of users. If an attacker learns more about the distribution of users in the anonymous region, it can be known that u_2 and u_4 are inaccessible in the time interval Δt_1. u_3 and u_4 are inaccessible in the time interval Δt_2. We can know from the intersection of three queries that only the initial user u sending the continuous queries is accessible, so the attacker knows that the real querying user is u. And then the user's location privacy is compromised.

B. Location Privacy Enhancement Scheme Based on Regional Reachability

To solve the above Maximum movement boundary attack problem that cause by ignoring the temporal and spatial Relations of anonymous interval, we propose that the reachability test of the temporal and spatial relation in adjacent anonymous region should be done to all the historical users in the common user set.

1) Generalization of user history footprint

We need to generalize a single user's historical footprint into a grid region, in order to allow users to reflect the historical footprint of regional relations better, and more reasonable to measure the temporal and spatial relationship between regions. As shown in Fig. 2, all historical user footprints in the selected area are divided into corresponding grid areas. There is a set $\{u_i\}$ of users for any grid g_i, and a collection $\{f_i\}$ of footprints for any user u_i. Assuming a single mesh area is A and A≤ A_p (A_p is the maximum mesh area parameter), then, we can reasonably assume that for any grid g_i, the time it takes to move within the grid is t and $t < \varepsilon$ (ε is a tiny value), regardless of where the user u_i is in the grid.

2) Regional reachability detection

In this paper, it is proposed that the regional reachability test should be performed on the candidate region set and the user set in constructing the anonymous region and the common user set. In this way can we prevent the attacker from using the Maximum movement boundary attack to narrow the anonymous area, thereby narrowing the common user set. The regional reachability is defined as:

We assume a pair of regions g_i and g_j, if $\exists u_k \in g_i \cap g_j \wedge u_k.f_i.t - u_k.f_j.t \leq \rho * \Delta t$($\rho$ is the time interval coefficient), Δt is the time interval between adjacent queries. Then we say that the regions g_i and g_j are reachable in the time interval Δt, and $\forall u_m \in g_i \cap g_j$ are reachable in the time intervalΔt.Then we extend this to the area reachability detection in consecutive queries. We suppose the adjacent query as $Q_i < u, t_i, l_i(x, y), k_i >$ and $Q_{i+1} < u, t_{i+1}, l_{i+1}(x, y), k_{i+1} >$, G^i is the anonymous region grid set of the ith query, G_c^{i+1} is the candidate region set of the (i + 1) th query. Then $\forall g_n \mid g_n \in G_c^{i+1}$, if $\exists g_m \mid g_m \in G^i$, and g_n to g_m are reachable, then $\forall u_k \in g_n \cap g_m$, u_k is reachable in adjacent queries, and is able to be the user in the common user set.

Based on the above, we design an anonymous region construction algorithm based on region accessibility. We assume that the user's continuous queries are $Q = \{Q_1, Q_2, \dots, Q_n\}$, $Q_i < u, t_i, l_i(x, y), k_i >$. First, we generate the candidate anonymous region G_c^i (dividing the grid) and the candidate common user set U_c ($|U_c| > k$), according to the existing anonymous region construction algorithm. We do a regional reachability check on all queries for each user u_i in U_c, excluding unreachable areas and users, forming the final anonymous area $A(A > A_{min})$ and common user set $U(|U| \geq k)$. Specific algorithm is as follows:

Algorithm 1: Privacy enhancement based on region reachability

Input: Queries sequence $Q = \{Q_1, Q_2, \dots, Q_n\}$;
Output:The final anonymous area $A = \{A_1, A_2, \dots, A_n\}$, *Common user set* U;

```
for each $Q_i \in Q$ do
get $G_c^i = dividetogrid(A_c^i)$ // According to the existing scheme, construct the candidate anonymous region and divide the grid
$G_c = \{G_c^1, G_c^2, \dots, G_c^n\}$
$U_c \leftarrow |G_c^1| \cap |G_c^2| \dots |G_c^n|$ // Obtain the candidate common set of users
for each $u_i \in U_c$
  for each $G_c^j \in G_c$
    If time($grid(u_i, G_c^{j+1}), grid(u_i, G_c^j)) > \rho * |Q_{j+1}.t - Q_j.t|$ do
            $G_c^j = G_c^j - grid(u_i, G_c^j)$; // Exclude the user's grid
        $U_c = U_c - u_i$; // Exclude the user
        if $|U_c| < k \wedge G_c^j < A_{min}$ do //Common user sets do not meet privacy queries
          Exit;
        end if
    end if
   end for
end for
return $U \leftarrow U_c$ A$\leftarrow G$;
```

IV. EXPERIMENTAL SIMULATION

In order to illustrate the security threats and the feasibility of this scheme, we design different simulation experiments in this paper. From the aspects of whether the common user set can resist the Maximum movement boundary attack and the computational cost of the algorithm, to point out the problems of current scheme and prove the effectiveness and practicability of the scheme of this paper.

All algorithms are implemented in Java programming language, and the experiments are performed in a PC with a Intel Core i7, 1.7GHz processor, 8GB RAM and Windows 7 operation system. The experimental data were provided by the GeoLife[10] project of Microsoft Research Asia. The data collected GPS trajectory of 182 volunteers during about 5 years, including walking, cycling, driving and other transportation mode. We choose about 2000 users' historical trajectory records within the 5th Ring of Beijing (about 20km * 20km) as the experimental data. As shown in Table 1.

TABLE I. GEOLIFE DATASET DATA COMPOSITION

Transportation mode	Distance (km)	Duration (hour)
Walk	10,123	5,460
Bike	6,495	2,410
Bus	20,281	1,507
Car & taxi	32,866	2,384
Train	36253	745
Airplane	24,789	40
Other	9,493	404
Total	140,304	12,953

A. *Algorithm Security Analysis*

In order to observe the security of this design, this section designed a set of contrast experiments. Through the implementation of the L2P2 proposal presented at the 2012 INFOCOM meeting, the accessibility of its common user set was observed. This experiment sets the user to do 20 consecutive queries for privacy queries, k value interval [10, 40], time interval coefficient $\rho = 1.5$, mesh area parameter $A_p = A_{min}$. Then do statistics on the results before and after reachability test of 5, 10, 15, 20 queries, and come to the following experimental results. As shown in Figure 3, as the number of consecutive queries increases, the number of users in the common user set of L2P2 scheme when it is accessible and inaccessible are decreasing. The reason is that with the growth of the line and the increase in the number of queries, finding the intersection of the original common user set and the current user set makes the number of users declining, but the figure shows, the number of users when it is reachable is less than before. This is because the common user set constructed by the L2P2 scheme contains time-unreachable users. When the reachability detection is introduced, these users will not become parts of the common user set and will be excluded from the common user set.

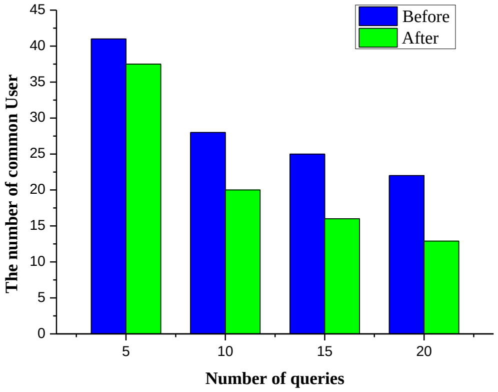

Figure 2. The number of common user sets for different frequency queries

As shown in Fig. 4, the ratio of unreachable users to common users in the L2P2 scheme is shown. With the increase in the number of queries, the proportion of unreachable users is growing, and the part of the user will not resist the Maximum movement boundary attack. Therefore, the Maximum movement boundary attack will be able to exclude more and more users and the user's privacy level will continue to decline. So the L2P2 scheme will become increasingly insecure with the increase in the number of queries. In this paper, we design a scheme which is more secure than the L2P2 scheme, because it considers the accessibility problem and eliminates the security risks of the unreachable users in the stage of constructing the common user set.

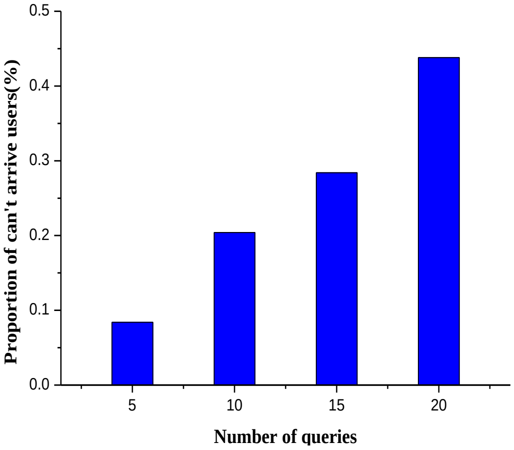

Figure 3. Unreachable users account for the proportion of common users

B. *Practicality of the Algorithm*

We also calculated the computational cost of the reachability detection algorithm for different frequency of query (n = 10, 15, 20) and different privacy queries (k = 20, 40, 60). In order to illustrate the introduction of accessibility testing does not cause too much computing cost on the anonymous process. It can be seen from the data in Table 2 that the calculation time is only 14.62ms in the case of 20 consecutive queries and privacy query of 60, so it will not burden the whole anonymous process and will not affect the service quality of users. In addition, this algorithm is based on the historical footprint information, and does not need to use map API for time accessibility judgment, there is no additional communication cost.

TABLE II. CALCULATED OVERHEADS /MS FOR DIFFERENT FREQUENCY OF QUERIES AND PRIVACY QUERY

Number of query k	10	15	20
20	3.15	5.32	6.14
40	5.66	7.53	8.57
60	6.87	9.98	14.62

V. SUMMARY

This paper firstly points out that the existing k-anonymous scheme based on historical footprints ignores the reachability problem. The analysis shows that the existing scheme can't resist the Maximum movement boundary attack. It is easy to narrow the anonymous region and reduce the common user set which leads to the user's privacy protection level down. And the experiment was carried out to validate the analysis. Therefore, this paper proposes that the time-reachability test should be performed on the candidate anonymous region and the candidate common user set before constructing the anonymous region. And a specific algorithm is designed to filter the candidate anonymous region and candidate common user set. Through analysis and experimental verification, this algorithm can effectively resist the Maximum movement boundary attack without increasing the extra burden of the system and guarantee the privacy queries of the users.

REFERENCES

[1] L. Sweeney. "Achieving k-anonymity privacy protection using generalization and suppression," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, pp. 571-588, 2002.

[2] M. Gruteser, D. Grunwald, "Anonymous usage of location-based services through spatial and temporal cloaking," Proceedings of the 1st international conference on Mobile systems, applications and services. ACM, pp. 31-42, 2003.

[3] M. F. Mokbel, C. Y. Chow and W.G. Aref, "The new Casper: query processing for location services without compromising privacy," Proceedings of the 32nd international conference on Very large data bases. VLDB Endowment, pp. 763-774, 2006.

[4] T. Xu, Y. Cai, "Exploring historical location data for anonymity preservation in location-based services," INFOCOM 2008. The 27th Conference on Computer Communications. IEEE. 2008.

[5] T. Xu, Y. Cai, "Feeling-based location privacy protection for location-based services," Proceedings of the 16th ACM conference on Computer and communications security. ACM, pp. 348-357, 2009.

[6] Y. Wang, D. Xu and X. He, et al, "L2P2: Location-aware location privacy protection for location-based services," INFOCOM, 2012 Proceedings IEEE, pp. 1996-2004, 2012.

[7] X. Li, L. Deng and S. Gao, et al, "A demand-aware location privacy protection scheme in continuous location-based services," International Conference on Connected Vehicles and Expo. IEEE, pp. 137-150, 2014.

[8] H. Liu, X.H. Li and E.M. Wang, et al, "User privacy enhancement method based on false location under continuous service query," Journal of Communications, vol. 37, 2016.

[9] M. Wernke, P. Skvortsov and F. Dürr, et al, "A classification of location privacy attacks and approaches," Personal and Ubiquitous Computing, vol. 18, pp. 163-175, 2014.

[10] Y. Zheng, Y. Chen and X. Xie, et al, "GeoLife2. 0: a location-based social networking service," 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware. IEEE, pp. 357-358, 2009.