Location Privacy Protection for Preventing Replay Attack under Road-Network Constraints

Lan Sun, Ying-jie Wu, Zhao Luo, Yi-lei Wang
College of Mathematics and Computer Science
Fuzhou University
Fuzhou, 350116, P. R. China
lsun@fzu.edu.cn

ABSTRACT: *Within the field of data privacy preserving research, privacy preserving Location-based Services for mobile users in road networks is a topic of great concern in recent years. However, most previous works on location privacy protection in road networks adopt spatial cloaking techniques, which may easily suffer replay attack. In addition, existing solutions for resisting replay attack generally add some randomized mechanisms into the cloaking process, which may reduce the query efficiency. In this paper, we firstly propose a group-division method to prevent replay attack, and then present two strategies based on this method to form the cloaking group of road segments. The experiment results on real datasets of road networks demonstrate the effectiveness and feasibility of our solutions*

Keywords: Location-based Services, Location Privacy Protection, Road Network, Replay Attack

Received: 18 May 2015, Revised 25 June 2015, Accepted 3 July 2015

1. Introduction

With the development of wireless network technology and communication technology, Location-Based Service (LBS) emerges in the right moment. It makes the queries of mobile users anytime and anywhere become possible. However, when mobile users enjoy the convenience of this kind of service, they need to send their accurate location information to the server, which will make their personal privacy suffer attack. How to ensure the high quality for service and do not reveal the location privacy of users has become a hot spot in the database field.

Recently, several techniques have been proposed to protect location privacy of mobile users [1-14]. Most of them are based on location *k*-anonymous model [1]. In Euclidean space, the most commonly used method to satisfy the location *k*-anonymous is the spatial cloaking technique [1-7, 9], whose main idea is to obscure the users' accurate locations into a closed spatial regions, named cloaking area. The core of this method is how to construct the cloaking area containing the user's original location. Different from the Euclidean space, in mobile networks, such as road networks, the user's moving direction is constrained by the influence of the network topology, but not arbitrary. However, the cloaking technologies in Euclidean space may leak of theusers' location privacy if they are directly used in road-network [10, 11, 13]. Therefore, some techniques that substitute cloaked road segment set for cloaking area are proposed to particularly protect the users' location privacy in road-network [8, 10-14].

road segment set for cloaking area are proposed to particularly protect the users' location privacy in road-network [8, 10-14].

Two anonymous algorithms-PSNN and PSRN for Nearest Neighbor Query and Range Query in road networks are proposed in [8]. However, as the two algorithms are based on the privacy-aware technology which is used in Euclidean space, the user location privacy still can be compromised. Li et al.[10] designs a hierarchical structure under road-network constraints to construct the location cloaking area. However, such a static hierarchical structure has a deterministic property for its cloaked area, so it is vulnerable to suffer from a reverse engineering attack [13], which is called a replay attack. The other work presents an anonymous algorithm based on edge sorting in [12], but as a result of using the fixed privacy requirement k for all users in road networks, it cannot meet the personalized privacy requirement of different users. In [11], a new location anonymization model, XStar, is proposed. And then, a general framework of location privacy protection is designed, which not only takes the query execution cost and communication cost into consideration, but also supports the personalized privacy requirements of mobile users. But the underlying basic star structure degrades the query processing efficiency and the shared execution has not been fully utilized[13]. In order to balance the query cost and service quality, a new query cost evaluation function is designed in [13]. Two greedy algorithms, pure greedy and randomized greedy are introduced in [13], whose aims are to form the cloaked segment set. The former algorithm fully considers all the segments in geographical proximity, and constructs the cloaked segment set by network expansion which guarantees the query quality and reduces the query cost. However, it is still vulnerable to suffer from a replay attack. In order to avoid this problem, in the randomized greedy algorithm, a random factor is added into the basis of pure greedy. Though the randomized solution can control the probability of replay attacks to some extent, it will increase the query cost in the meantime.

The studies described above can effectively protect the mobile users' location privacy under the assumption that the service provider is always reliable while attackers know nothing about cloaking algorithm. However, the adversary often has some prior knowledge about anonymous algorithm and the cloaked segment set, so the replay attacks will occur with high probability. In this paper, we propose two strategies of constructing cloaked segment set under road-network constraints, whose aims are to prevent replay attack with high query efficiency.

The rest of this paper is organized as follows. The preliminary is described in Section 2. In Section 3 we present our solution to prevent replay attacks. The experiment results are showed in Section 4 and in Section 5 we conclude our works.

2. Preliminary

2.1 Road Network Model

Without loss of generality, in this paper, we model a road network as an un-directed graph $G = (V, E, W)$, where V represents a set of road junctions and E represents a set of links between the roads, while $w_{ij} \in W$ indicates the number of users in edge e_{ij}.

Definition 1 (Segment) Road segment s is a set of edges, $s = (v_0v_1, v_1v_2, ..., v_{L-1}v_L)$, where $\{v_i\}_{i=0}^{L}$ are distinguishing, and when $i = 0$ or L, the degrees of the nodes greater than or equal to 3, but for others, the degrees are 2. [11]

It is quite clear that any edge is either a segment itself, or belongs to a unique segment. So we assume that each mobile user is moving along certain road segment.

2.2 System Model

In this paper, we choose the centralized system model, the same architecture as that in many existing works. The model is composed of three parts-users, an anonymizing proxy and a location server, as shown in Figure 1. In this architecture, it is assumed that the communication between the users and the anonymizing proxy is reliable, while the communication between the anonymizing proxy and the location server is unreliable. The anonymizing proxy keeps the original location of users all the time. When a user issues a query, the whole process is executed as follows: (1)the query containing an accurate location is sent to the anonymizing proxy; and then (2)the anonymizing proxy anonymize the original location of the user into a set of segments, named cloaked set, and then the anonymizing proxy sends the cloaked set to the location server; (3)the location server executes the query according to the cloaked set and generates candidate results; (4) the anonymizing proxy receives the candidate results from the location server, and sends the refined results to the user after filtering the wrong result. During the whole process, the actual location of the user is unknown to the location server, so the location privacy is protected.

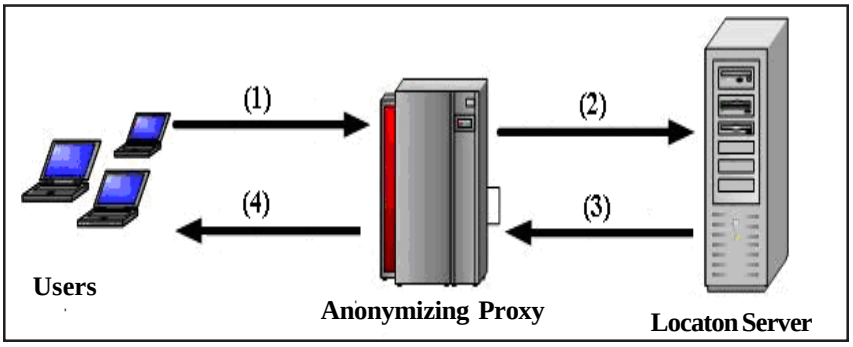


Figure 1. The Centralized System

2.3 Replay Attack

The replay attack may take place in the situation when the adversary has the priory knowledge of the cloaked set and the anonymization algorithm. For each segment *s* belonging to cloaked set *S*, by re-running the anonymization algorithm with the assumption that *s* is the original location, the adversary can estimate the likelihood of *s* to generate the cloaked set *S*, denoted by *like* (*s*, *S*).

The whole process of replay attack is as following: for each segment *s* in cloaked set *S*, (1) run the anonymization algorithm A(.) with *s* and generate a cloaked set of segments, *S′*; (2) by comparing the two sets, *S* and *S′*, compute the likelihood like (*s*, *S*) = $||S' \cap S| / |S|$; then (3) select the segment *s** that leads to the largest likelihood value as the original location, namely,

$$s^* = argmax_s \; like\,(s, S) \qquad (1)$$

In some situation, there may be two or more segments leading to the largest likelihood, the adversary can only choose one from them randomly as the original location.

If the segment which the adversary chooses by launching replay attack is the true location where the query originates from, the replay attack succeeds.

3. Strategies Against Replay Attack

On the basis of description and explanation of replay attack executing process, it is quite clear that an ideal situation for preventing replay attack is that each segment *s* in cloaked set *S* has the same likelihood as the original location. In this case, the road segments in the cloaked set are indistinguishable for adversaries. Therefore, attackers cannot know anything about the road segments in which the query is generated and the replay attack is thus avoided.

Definition 2 (Anonymous equivalence). Given a anonymization algorithm A(.), for two road segments in road networks, s_1 and s_2, where two queries are generated respectively, running the anonymization algorithm with s_1 and s_2 would generate two cloaked sets, $A(s_1)$ and $A(s_2)$. If $A(s_1)$ equals to $A(s_2)$, we will say the segment s_1 is anonymous equivalent to segment s_2 in the mean of algorithm A(.), representing as $s_1 \sim s_2(A(.))$.

Based on the definition of anonymous equivalence, two road segments which are anonymous equivalence must have the same likehood as the original location. The main idea of the solution in this paper is to find a cloaking algorithm *A*(.), generating a cloaked set *A*(*s*) with respect to a segment *s*; For each segment *s′* in *A*(*s*), *s′* is anonymous equivalent to *s*, that is, $s_1 \sim s_2(A(.))$. For two cloaked sets, $A(s_1)$ and $A(s_2)$, if $A(s_1) \neq A(s_2)$, then the intersection of two different cloaked sets is empty. Obviously, if we can present such an algorithm *A*(.), the replay attack will be prevented, because each segment in the same cloaked set has the same likelihood as the original location.

A cloaked set consists of a series of road segments, so we can divide the all segments into several "buckets". If the segments in each bucket not only satisfy users' privacy requirements, but also are anonymous equivalence, the segments in the bucket can be used as a cloaked set and can prevent replay attack. In our study, we represent the user's privacy requirement as <*k*, *l*>, meaning there must be

at least *k* users and *l* road segments in a cloaked set. It is noticeable that different users in the same bucket usually have different privacy requirements. So each bucket must meet the maximum of different privacy demands, represent as <*kmax*, *lmax*>. That is to say each bucket contains at least *kmax* users and *lmax* road segments.

3.1 Sequential Scan Grouping

In this paper, we use two strategies to form the buckets satisfying the privacy requirement. The first one is sequential scan. We sequentially scan the sorted segment sequence and select the first segment as a temporary group. Then test whether the segments in the temporary group can form a cloaked set satisfying the user's privacy requirement. If they can, we store all the segments in the temporary group as a new bucket and clear the temporary group; otherwise, we scan the next segment in the segment sequence and repeat this step until all the segments are scanned. If the segments in the temporary group cannot form a cloaked set meeting privacy requirement at the end of the scanning processing, we put all the segments of the temporary group into the bucket built lastly. The algorithm is described as follows.

```
Algorithm 1: Sequential Scan Grouping
//T: temporary group//
S: the sorted segment sequence according to the numbers of the
users residing in
//ResSet: a set preserving the decomposition result of segments
 (1) T ← the first segment in S
(2) while (sequential scan is not finished )
 a. while ( T cannot satisfy the privacy requirements of all users
in T)
 b.   if (sequential scan is not finished)
       T← the next segment in S
  c.   else
          break;
 d. end while
 e. if (T can satisfy the privacy requirements of all users in T)
ResSet ← T
 f. else
  merge the T with the last element in ResSet
 g. T ← empty
(3) end while
```

3.2 Binary Grouping

It is obvious that the last bucket maybe has certain redundant in sequential scan. We propose the other top-down strategy which still starts from a sorted segment sequence for all road segments. The implementation process of this strategy is as follows. Firstly, if the initial sequence can be divided into two parts which both can meet the maximum privacy requirement of users residing in segments in each part, it is divided into two groups. The dividing process is repeated recursively for both of the two groups until a group cannot be divided into two subgroups satisfying the maximum privacy requirement any more. In the dividing process, if a group cannot be divided into two subgroups satisfying the maximum privacy requirement any more. We call the group a bucket. When a request happens in one segment, we use the bucket containing the segment as the cloaked set. Because each bucket has at least *kmax* users and *lmax* road segments, so it can guarantee the user's location privacy. Specific algorithm is described below.

Notice that both strategies start from a segment sequence sorted by the number of users residing in. The reason lies in that it can make segments having similar number of users adjacent in the sequence. Based on the sorted segment sequence, the two strategies not only can prevent replay attack, but also can resist the weight-based attack [14, 17].

```
Algorithm 2: Binary Grouping
//Q: a queue consisted of a series of segment sets
 //S: segment set sorted by the number of users residing in each segment
//ResSet: preserves the decomposition result of all segments
 (1) Q ← S
(2) while(Q is not empty)
a.  S0 ← pop the first element of Q
b.  if(S0 can be divided into two subsets S1 and S2 satisfying the privacy
                                        requirement) divided S0 into S1 and S2,
half and half push the S1 and S2 into            Q
c.  else
    ResSet ← S0
(3)end while
```

4. Experiments

4.1 Experiment Settings

The simulated experiment is implemented in Visual C++ 6.0 under Windows XP SP3 and run in a machine with Intel(R) Core(TM)2 Duo CPU T6600 @2.2GHz, 2.2GHz CPU and 2GB main memory. The default values and evaluation ranges of each parameter are showed in Table 1.

Our experiments were performed over real road map of Oldenburg city in Germany, which contains 6105 nodes and 7035 segments, and the average segment length is 184m. We used the network-based generator of moving objects [16] to generate the location dataset. In order to express conveniently in the following figures, BG and SG are used to denote the two algorithms used in this paper, binary grouping and sequential scan grouping respectively, while BA represents the balanced adopt algorithm in [17], and PG and RG represent the pure greedy algorithm and the randomized greedy algorithm in [13] respectively.

Without loss of generality, all experiments randomly generate 1000 query requests and take the average result of these query requests as the final experimental result.

We compare the effect of the algorithms by the service response time which is consisted of anonymization time and query processing time. Obviously, the algorithm with shorter service response time is more effective. Because the service response time is affected by many factors, so we perform the simulation with four different aspects, including the privacy requirement, the number of users, the number of target objects and query level *K* of *K*-NN query.

4.2 Anonymization Time

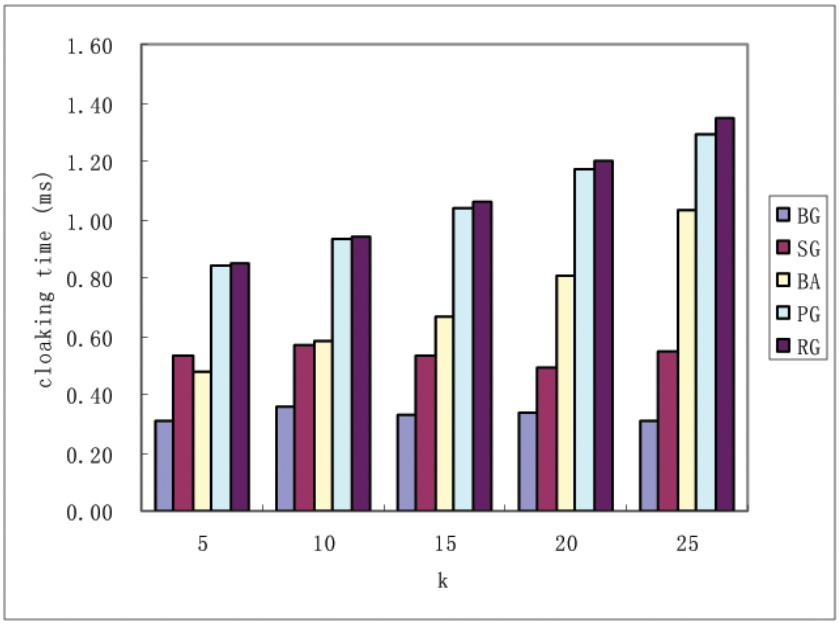
The change of anonymization time under different values of impact factors is shown in Figure 2. From Figure 2(a), we can find that, as the privacy requirement level increases, BA, PG and RG spend more anonymization time than the two algorithms, BG and SG, in this paper. The reason is that BA, PG and RG generate cloaked set by network expansion. When the privacy requirement increases, they need add more segments into the cloaked set. However, the two algorithms of this paper, BG and SG, use dividing the segment sequence into a series of buckets to form cloaked sets. When the anonymity level *k* increases, the buckets formed by the dividing may be large, so the dividing process of BG can terminate earlier. And for SG, it must scan all the sequence to form the cloaked set, so the anonymization time remains the same with different *k*. It is easy to find in Figure 2(b) - 2(d) that the anonymization time does not change when the number of users in whole road networks, the number of target objects and query level *K* increase. Figure 2 also tells us BG spends the least anonymization time in all the five algorithms. It really outperform the other four algorithms.

4.3 Query Processing Time

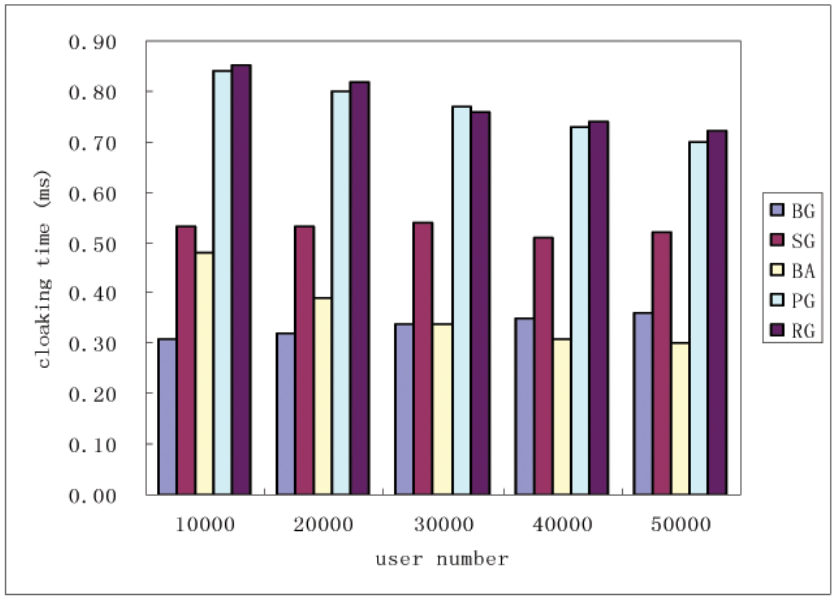
The change of query processing time under different values of impact factors is shown in Figure 3. It is easy to find that, as the privacy requirement *k* and query level *K* increase, the query processing time increases, while as the number of users and target objects increases, the query processing time decreases. When privacy requirement increases, there are more segments

Parameter	Default Value	Evaluation Range
Anonymity level(*k*)	100	100 – 500
Number of users	10000	10000 – 50000
Number of target objects	200	200 – 1000
Query level(*K*)	3	1 – 10

Table 1. Parameter Settings

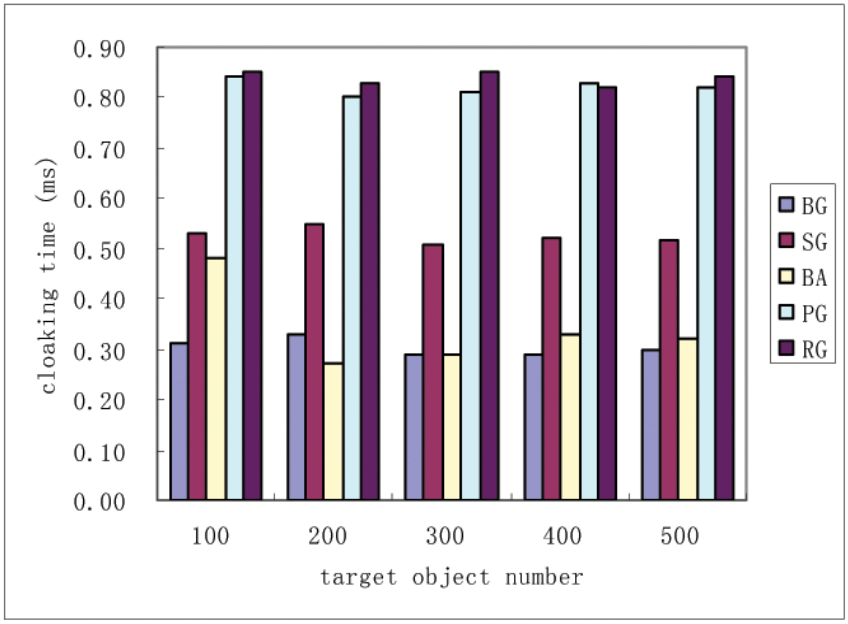


(a) Change of cloaking time under different *k*

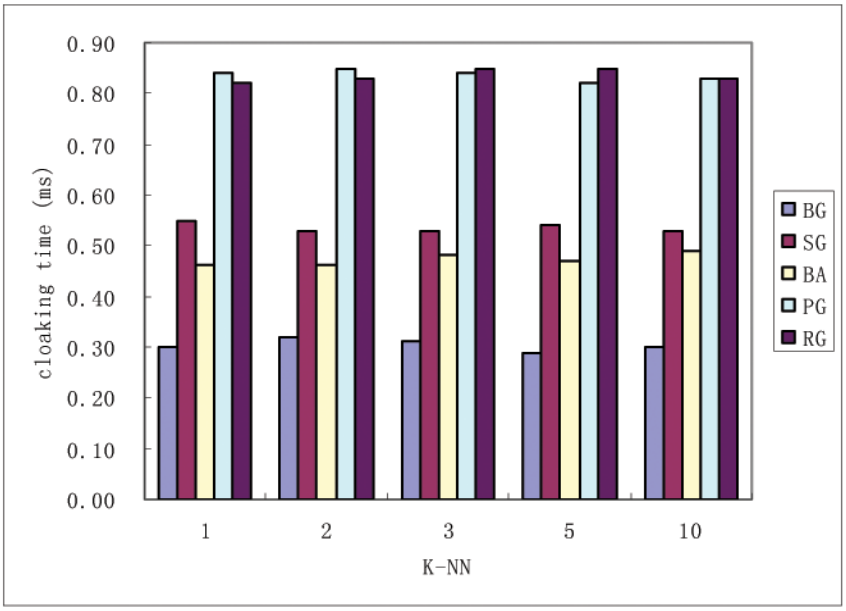


(b) Change of cloaking time under different user's number

in the cloaked set which leads to the query processing time increasing. And with the increasing of query level, the location server must search a larger region to complete the query process. While as the user number increases, the average number of users residing in each segment increases too. It means that there are fewer segments in cloaked set for the certain privacy requirement, so the query processing time is less. Similarly, as target objects increases, the location server only need to search a smaller region to find the *K* nearest targets, so the query processing time decreases. BG and SG in this paper expend more query time than other three algorithms. But, on the whole, the total service response time of BG and SG still has certain advantages.

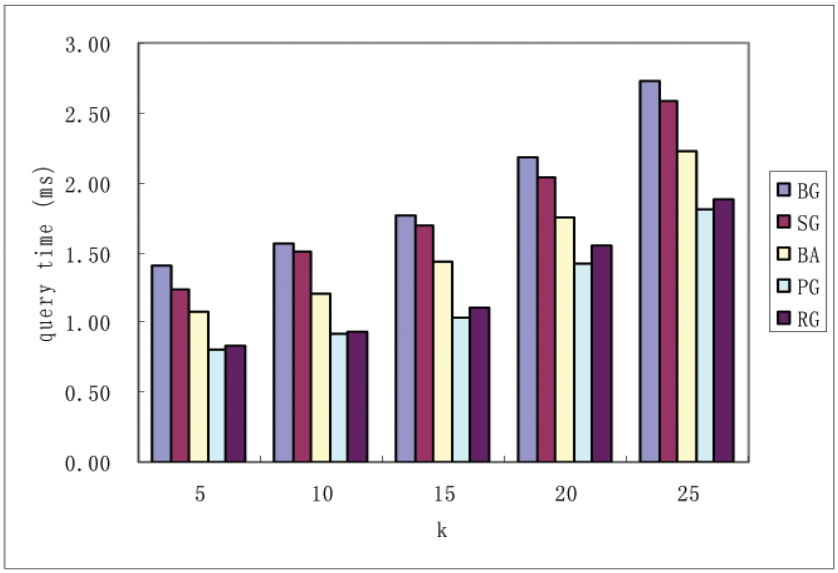


(c) Change of cloaking time under different target object's number

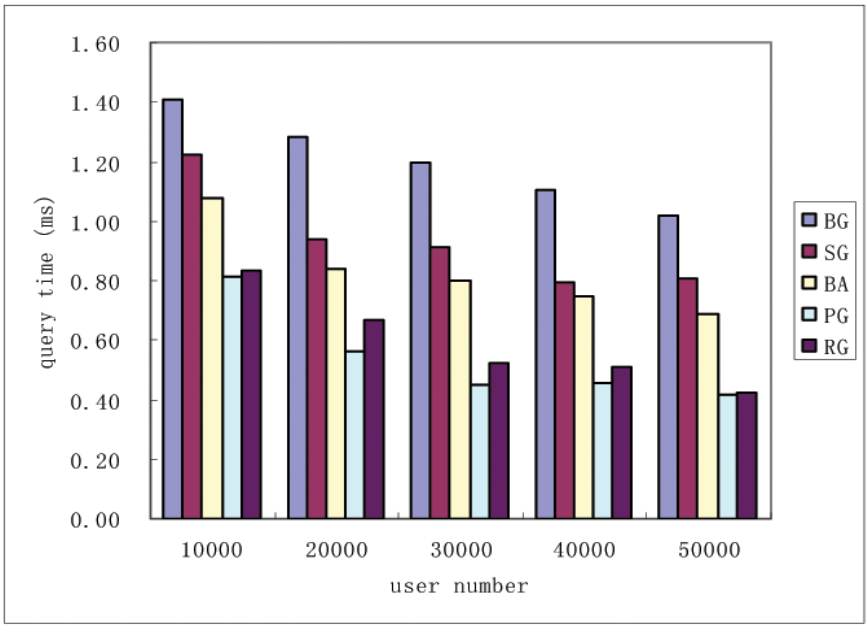


(d) Change of cloaking time under different query level

Figure 2. The Evaluation of Anonymization Time



(a) Change of query processing time under different *k*



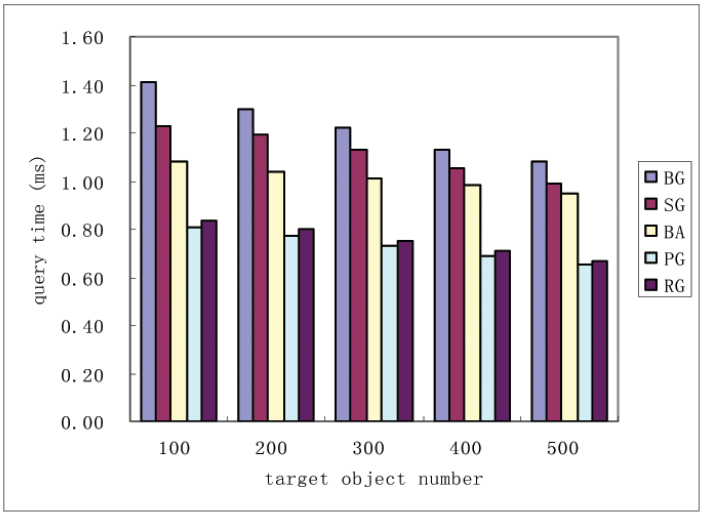
(b) Change of query processing time under different user's number

5. Conclusion

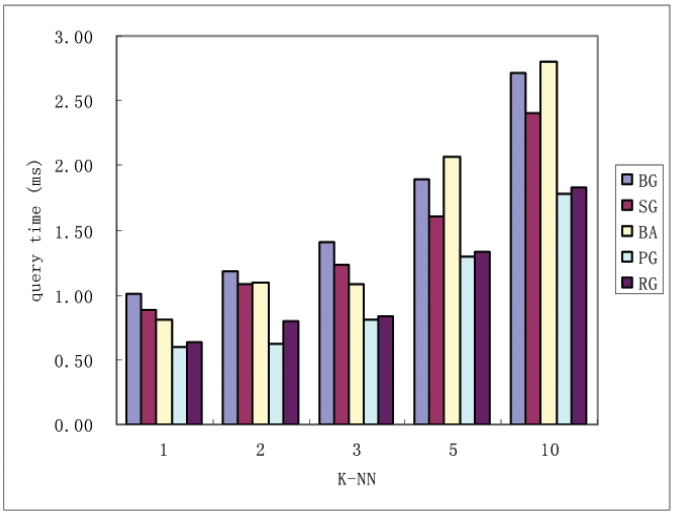
In this paper, we present two algorithms based on dividing to form the cloaked sets. Both algorithms can prevent replay attack to some extent, while guaranteeing the well distribution of weight in the cloaked set. The experiment results show that our algorithms are effective and feasible.

Acknowledgments

This study has been funded by the National Natural Science Foundation of China under Grant No. 61300026 and by the Development Foundation of Fuzhou University under Grant No. 2012-XQ-27.



(c) Change of query processing time under different target object's number



(d) Change of query processing time under different query level

Figure 3. The evaluation of Query Processing Time

References

[1] Gruteser, M., Grunwald, D. (2003). Anonymous Usage of Location-based Services through Spatial and Temporal Cloaking. *In*: Proc. 1st International Conference on Mobile Systems, Applications and Services, Scan Francisco, USA: *ACM Press*, 31–2.

[2] Papadias, D., Zhang, J., Mamoulis, N., et al. (2003). Query Processing in Spatial Network Databases. *In*: Proc. 29*th* International Conference on Very Large Data Bases(VLDB), Berlin, Germany: *VLDB Endowment*, 29, 802-813.

[3] Gedik, B., Liu, L. (2005). Location Privacy in Mobile Systems: A Personalized Anonymization Model. *In*: Proceedings 25*th* International Conference on Distributed Computing Systems. Columbus, USA: *IEEE Press*, 620-629.

[4] Mokbel, M. F., Chow, C. Y., Aref W. G. (2006). The New Casper: Query Processing for Location Services without Compromising Privacy. *In*: Proceedings 32nd International Conference on Very Large Data Bases (VLDB), Seoul, Korea: *VLDB Endowment*, 763–774.

[5] Cheng, R., Zhang, Y., Bertino E., et al. (2006). Preserving User Location Privacy in Mobile Data Management Infrastructures. *In*: Proceedings 6th International Privacy Enhancing Technologies Symposium (PET), Cambridge, UK: *Springer Berlin Heidelberg*, 393-412.

[6] Kalnis, P., Ghinita, G., Mouratidis, K., et al. (2007). Preventing Location-based Identity Inference in Anonymous Spatial Queries, *Knowledge and Data Engineering, IEEE Transactions on*, 19 (12), 1719–1733.

[7] Chow, C. Y., Mokbel, M. F. (2007). Enabling Private Continuous Queries For Revealed User Locations. *In*: Proceedings 10th International Symposium on Spatial and Temporal Databases (SSTD), Boston, MA, USA: *Springer Berlin Heidelberg*, 258-275.

[8] Ku, K. S., Zimmermann, R., Peng, W. C., et al. (2007). Privacy Protected Query Processing on Spatial Networks. *In*: Proceedings 23rd IEEE International Conference, Istanbul, Turkey: *IEEE Press*, 215-220.

[9] Bamba, B., Liu, L., Pesti, P., et al. (2008). Supporting Anonymous Location Queries in Mobile Environments with PrivacyGrid. *In*: Proceedings 17th international conference on World Wide Web, Beijing, China: *ACM Press*, 237-246.

[10] Li, P. Y., Peng, W. C., Wang, T. W., et al. (2008). A Cloaking Algorithm Based on Spatial Networks for Location Privacy. *In*: Proc. Sensor Networks, Ubiquitous and Trustworthy Computing (SUTC'08), IEEE International Conference on, Taichung, Taiwan: *IEEE Press*, 90-97.

[11] Wang, T., Liu, L. (2009). Privacy-Aware Mobile Services over Road Networks. *In*: Proceedings of the VLDB Endowment, (2) 1, 1042-1053, August.

[12] Mouratidis, K., Yiu, M. L. (2010). Anonymous Query Processing in Road Networks. *Knowledge and Data Engineering, IEEE Transactions on*, 22 (1), 2–15, January.

[13] Chow, C. Y., Mokbel, M. F., Liu, X. (2011). Query-Aware Location Anonymization for Road networks. *GeoInformatica*, 15 (3), 571-607.

[14] Sun Lan., Luo Zhao., Wu Yjingjie., et al. (2012). An Algorithm for Protecting Location Privacy in Road Network. *Journal of Shandong University (Engineering Science)*, 42(5), 96-101.

[15] Xiao, Z., Meng, X., Xu, J. (2007). Quality-Aware Privacy Protection for Location-Based Services. *In*: Proc. the International Conference on Database Systems for Advanced Applications (DASFAA'07), Bangkok, Thailand: *Springer Berlin Heidelberg*, 434-446.

[16] Brinkhoff, T. (2002). A Framework for Generating Network-Based Moving Objects. GeoInformatica, 6 (2), 153-180.

[17] Hao, Zhou.,Wang, Xiaodong., Sisi, Zhong (2012). A Weight-based Attack Model in Road Networks. *In*: Proceedings Industrial Control and Electronics Engineering (ICICEE), Xi'an, China: 1591-1594.