Continuous Anonymity with Users in the Same Direction in Road Networks

Yu Lili[1], Zhang Lei[1] ,Su Xiaoguang[1] Li Jing[1] ,Zhang Xu[1] and Wang Zhe[1]

1 College of Information and Electronic Technology, Jiamusi University Jiamusi, China

Abstract. Compared with the Euclidean space, road network is restricted by its direction in traveling, velocity and some other attribute profiles. So the algorithms that designed for the Euclidean space are usually invalid and difficult to provide privacy protection services. In order to cope with this problem, we have proposed an algorithm to provide the service of collecting anonymous users that their directions in traveling similar with the initiator in the road networks. In this algorithm, the shortest distance between multiple road segments is calculated, and then utilizes the distance to select the user who has the same direction in traveling with the initiator. Consequently, the problem of the discrepancy of the anonymous users in the routing that invalidates the location privacy protection is solved. At last, we had compared this algorithm with other similar algorithms, and through the results of the comparison and the cause of this phenomenon, we have concluded that this algorithm is better not only in the level of privacy protection, but in the performance of execution efficiency.

1 Introduction

Recently, with the development and popularization of mobile positioning technology, location-based services (LBSs) have been widely used in nearly all walks of life and have achieved good results in the applications [1]. However, as this service requires the user to provide real location information, and some time the location information may be the private information that users are reluctant to disclose to a certain extent. Thus, the problem of privacy protection in LBSs has brought about great concern of the researchers, and they had put forward a good deal of algorithms. Such as user collaboration [2], dummy location [3], probability indistinguishable [4] as well as profile attributes generalization [5] and so on.

However, as these algorithms were mainly designed to provide privacy protection service for users in Euclidean space, and the user's real living space is the environment of road network rather than Euclidean space, which made the failure of these algorithms in resisting attacks in this environment. Furthermore, as restrictions of different users in the road network are varied (such as the direction of traveling, velocity and some other attribute profiles), which leads the condition of privacy protection particularly seriousness. For this condition, several algorithms were proposed and used in the environment of road network, such as the mix - zone [6], the similar trajectory [7], the speed prediction [8], the generation of safety region [9] as well as trajectories collaboration [10]. However, these methods fail to effectively solve the problem of anonymous users of differences caused by mobile direction in the process of continuous query privacy protection of the road network. In order to solve this problem, based on the shortest path computation [11], this paper proposed a similar direction anonymous users finding algorithm short for (SDAA) to choose anonymous users. With this algorithm, the central server can select anonymous users by calculating the shortest distance of multiple road segments, then with the same direction of traveling to reduce the probability of being identified by the adversary. At last, this algorithm can solve the problem of anonymous users of differences caused by mobile direction in the routing in the road network.

2 The system architecture and the basic conception

2.1 The system architecture

In the process of privacy protection in SDAA, the shortest distance of multiple road segments has to be calculated, and utilizes the result to select anonymous users with the same direction. So the cost of calculation will be larger for just the device of users, accordingly a central server is needed and utilized the three layers system architecture. The architecture is shown in Figure 1. From this figure, we can see that, three different entities are proposed in this architecture and they are the mobile user, the central server and the location server respectively. For the mobile user, it is the user who initiates for location services and equipped with the device that can communicate with the central server as

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (http://creativecommons.org/licenses/by/4.0/).

well as the location server. The central server can be seen as a provider of privacy protection it collected the information from the mobile user and disposes it with others to reduce the probability of being identified by the adversary. The third entity is the location server it can be seen as the location service provider, and can provide service with the query sent by the mobile user. Furthermore, this entity was usually seen as an un-trusted entity, as it may be breached by the adversary or the jeopardize private information by commercial interests.

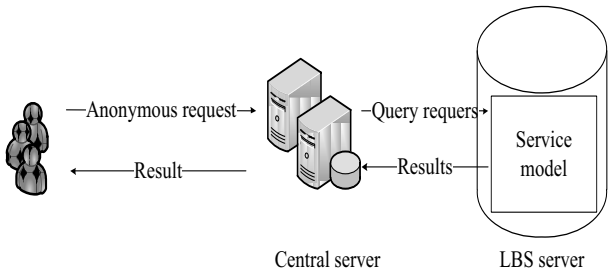

Figure 1 The architecture of centralized model

2.2 The privacy threat

In the environment of the road network, the process of user utilizes continuous query in the routing will produce a series of discrete locations, and these locations can be correlated and generated a location trajectory. As the trajectory may contain more spatio-temporal information than discrete locations, this trajectory will be even hazard to the privacy of the user. Conventional methods used in continuous query usually search for similar anonymous users in each location along the routing, and achieve the location generalization of the real user through the help of anonymous users. However, this kind of generalization may lead the differences about the direction of anonymous users movement during the continuous movement, and leads to the failure of continuous anonymity. For example: four users were generalized in the left region of the figure 2, and the user A can establish an anonymous group with B, C and D. However, as this user moving to the subsequent position and launches another query in the routing, she will establish another anonymous group with E, F and G, due to the restriction of road network. As a result, the user A will be identified by the adversary, as anonymous users are changed and cause anonymous differences. Table 2. Formatting sections, subsections and subsubsections.

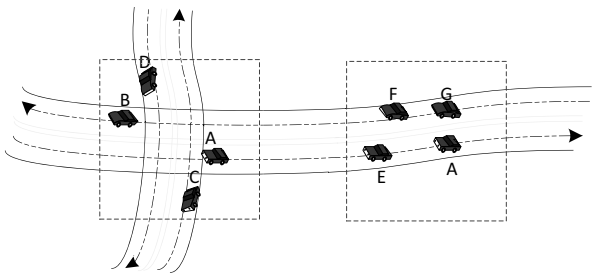

Figure 2 The difference of anonymity in road networks

In order to utilize this kind of failure, the adversary can design a simple method to analyse the location and identify the real user. Suppose the set of anonymous users can be denoted as $A=\{A_1,A_2,...A_k\}$, two arbitrary chosen sets about the anonymous users in the following locations of her routing can be denoted as $A'=\{A'_1,A'_2,...A'_k\}$ and $A''=\{A''_1,A''_2,...A''_k\}$, if $A\cap A'\cap A''=X$ and the sign X denotes the same user, it means the sequence that signed by X will be the real trajectory of a specified user. Accordingly, the adversary will get the trajectory and obtain the privacy.

2.3 The basic idea and conception

As the adversary can identify the real user by distinguishing the difference of anonymous users in continuous generalization region, and this operation can be achieved by the sign calculated from $A\cap A'\cap A''=X$. If an algorithm can protect the privacy, which means it must generalize the sign of each anonymous user in each generalization region. In order to achieve the sign generalization, the algorithm has to choose anonymous users that have the same direction of travelling, so that the set of anonymous users will be consistent and cause anonymous differences during the routing in the road network. Thus, based on this conception, the central server has to select anonymous users by the similar direction of the real user. In this paper, based on the calculation of the shortest distance of multiple road segments, a convenient method is proposed to select anonymous users in the same direction with the comparison of shortest distances. Then the privacy of the user in continuous query along the routing in the road network is protected.

3 The scheme of same direction users finding in multi road segments

Because of the direction of travelling is restricted in the routing, the privacy protection algorithm has to judge whether the selected anonymous users have the same direction, especially when the user is located in the road network. In order to convenient the process of estimating the mobile user's direction, the distance between each user can be utilized. Suppose two users have the same direction, the distance of them can be denoted as $d(A,B)$, if they have different direction the distance will be $d(B,A)$. As two distances $d(A,B)$ and $d(B,A)$ may be different from each other, we can utilize the comparison result of them to estimate whether they are running in the same direction. So we utilize the distance of multiple road segments to estimate the result and denoted it as

$$D=\min(d(A,B)) \tag{1}$$

Where $d(A,B)$ denoted the distance between two users A and B, min(.) means the shortest one among these results. Suppose two users at the crossroads of a road network are denoted in figure 3. Users are denoted as A and B, the terminal points of each road segment are T, V,

W and U respectively. Therefore, the shortest distance can be calculated from the comparison of the following results $d(A,W)+d(W,T)+d(T,B)$, $d(A,W)+d(W,U)+d(U,B)$, $d(V,A)+d(V,T)+d(T,B)$, $d(V,A)+d(V,U)+d(U,B)$. The distances such as $d(W,T)$ and $d(W,U)$ are the distances between two segments. Others are distances of users in this road segment. In order to alleviate the calculation, we confirm the distance in reverse direction as $+\infty$, but the distance in the same direction as above mentioned formulas

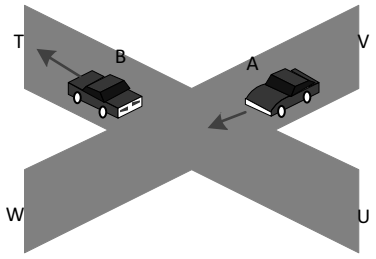

Figure 3 The crossover between two different road segments

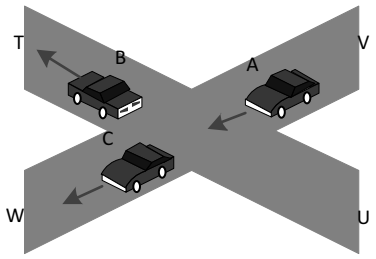

Figure 4 The distances between users in different directions

According to the above method, we can get the relation of distances calculated from the figure 4, and get the result that $d(A,C)<d(A,B)$. As the shortest distance can be compared by $\min(d(A,B))>d(V,W)>\min(d(A,C))$, if the user A wants to protect the privacy along her routing in road network shown in figure 4, the central server has to select the user C as the anonymous user to establish the anonymous group instead of user B. The process of how the central server selects the anonymous user is denoted in algorithm 1.

Algorithm 1 The collection of users in same direction

Input: The set of users in this region U， The anonymous value k

Output: The set of available anonymous users U_a

1) Calculates the distance $d_{current}$ of the user in current road segments；
2) for(i=1, i<=U.num, ++i)
3) Calculates the distance d between the real user and the anonymous user；
4) if($d <= d_{current}$)
5) add U_i into the set of U_a；
6) else
7) continue；
8) end if
9) if(U_a.num>=k)
10) break;
11) end if
12) end
13) output: U_a;

From the algorithm 1, we can see the process of selection the shortest distance in the line of 4-8, and the process of the chosen the enough number of anonymous users in 9-11. As this algorithm is designed to utilize the shortest distance to choose anonymous users in the same direction, and established the generalization group to achieve privacy protection the anonymous value is determined by the real user.

4 Security analyses

In order to verify the security of our proposed algorithm in this paper, the analysis will be initiated in the privacy protection capability and the correctness of the shortest distance two aspects. Firstly, the capability of privacy protection is determined by the uncertainty of the adversary successfully guesses the real user. As we have depicted in section 2, the adversary can utilize the calculated value of X to identify the trajectory of the user with $A\cap A'\cap A''=X$. From this equation, we will find that is the value of X is large enough or satisfied $X\geq k$, the success ratio of the adversary will be reduced to randomly guess, as the sign of the real user is generalized as no less than k. In the algorithm of SDAA, all anonymous users have the same direction of travelling with the real user, and this same direction will lead to the similar anonymous users in the next direction for the following query. It means that anonymous users are not changed, and there is no changing in the set of generalized users. As a result, the value of X is unchanged and satisfied $X\approx k$, so the probability of the adversary successfully guesses the real user will be dropped to 1/k, which is difficult to identify the real user.

For the correctness of the shortest distance, as the similar direction is calculated by the comparison of different distances, it is important to verify its correctness and confirms two users have the same direction when they have the shortest distance than others. Suppose the relationship of users is depicted in Figure 4, then we will have various values of distances denoted each user and we denoted in the follows.

$$d(A,B)=\min(d(A,B))=d(A,W)+d(W,T)+d(T,B) \quad (2)$$

$$d(A,C)=d(A,W)+d(W,C)+d(W,W) \quad (3)$$

If there exist $d(A,B)<d(A,C)\leq d_{current}$, then

$$d(A,W)+d(W,T)+d(T,B)<d(A,W)+d(W,C)+d(W,W) \quad (4)$$

According to the definition that $d(W,T)=+\infty$ and $d(W,W)=0$, we can have $d(A,B)>>d(A,C)$, which means the assumption is inaccuracy, and we can conclude that two users have the shortest distance must have the same direction. Therefore, through above analysis, it can be concluded that the algorithm proposed in this paper can guarantee the privacy of users in the process of

continuous anonymity under the road network environment.

5 Experimental verification

In order to further illustrate the advantages of the proposed algorithm in this paper, several experimental verifications are proposed for the comparison with other similar algorithms to verify the privacy protection capability and algorithm execution efficiency. The experiments are executed in the laptop with Intel core i5 processor and 4GB memory, and tested by using Matlab 2017Ra in Windows 7 operation system. At the same time, comparison algorithms involve the anonymous interval algorithm CORP [12], the collaborates trajectory algorithm LTPPM [13], similar tracks anonymous algorithm STG [5]. The comparison of the above algorithms will be carried out in three aspects: the anonymous success rate, the execution time of the algorithm and the identification rate of the adversary.

Figure 5 shows the variation of the anonymous success rate caused by the changing of continuous query time in the case of the anonymous value is 20. From this figure, we can see that all algorithms lead to a decrease in the success rate of anonymity as the number of queries increasing. Among these algorithms, the anonymous success ratio of SDAA is higher than other algorithms, as this algorithm selects anonymous users by their direction in road network, and the same direction with the real users can guarantee the anonymity in the process of continuous query along the routing of road network. For the other algorithms, as the CORP algorithm utilizes the precomputation method to select anonymous users it is harder to confirm that the uniform of anonymous users along the routing in each anonymous region, so its success ratio is a bit lower. STG utilizes the dummy similarly trajectories that generated by the central server to achieve generalization, which is suitably used in Euclidean space but seriously restricted in the road network, so its success ratio of anonymity is lower than CORP. At last, the success ratio of LTPPM is the lowest, this is because of this algorithm mainly utilizes the collaborative users who have the similar direction to anonymity the real user, but it is difficult to find enough collaborative users that have the similar direction, so its success ratio of anonymity is lower than all other algorithms.

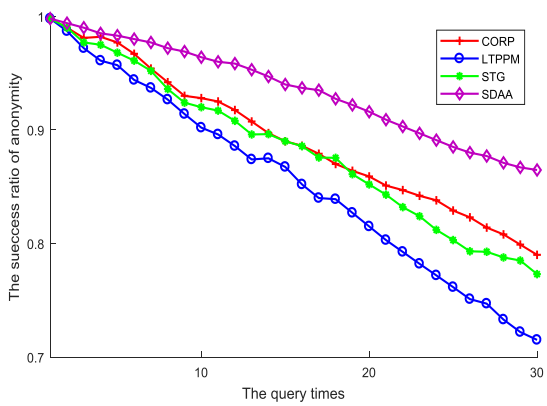

Figure 5 The success ratio of anonymity vs. query times

Figure 6 shows the success ratio of anonymity changed with the increasing of anonymous value. In this experiment we defined the continuous query time to 10, and anonymous is ascending to 30 gradually. It can be seen from the figure, all success ratios of algorithms are descending with the increasing of anonymous value, as the a larger anonymous value means a larger number of anonymous users have been selected, which leads no matter the central server or the user himself harder to establish the designed anonymous group. However, as SDAA mainly utilizes the central server to select anonymous users it performs better than algorithms mainly utilize collaborative users to achieve anonymity. Thus, its success ratio of anonymity is higher than the algorithm of collaborative users finding or the algorithm of precomputation or the algorithm of dummy similar trajectories, such as LTPPM, CORP and STG.

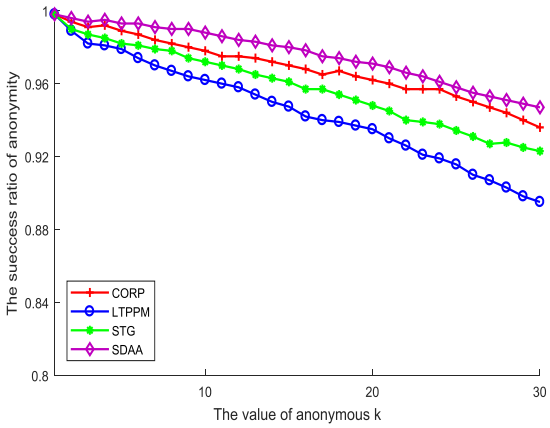

Figure 6 The success ratio of anonymity vs. anonymous value *k*

Figure 7 shows the variety of execution time of different algorithm along with the increasing of anonymous value. It can be seen from the figure, the execution time of all algorithms are ascending with the increasing of anonymous value, as each algorithm needs to find enough anonymous users according to the requirements of the anonymous value. The process of anonymous users finding is a time consuming process. Among these algorithms, the execution time of STG is the highest, as this algorithm needs to calculate each node of the location that will affect the similarity of the dummy trajectory, so its computational complexity is higher than other algorithms. LTPPM has to find collaborative users who can move in the same direction, but the discreteness of the collaborative users affects the process of anonymity, and its execution time is mainly consumed in this process, so this time is higher than CORP and SDAA. As the algorithm of CORP has to compute the potential interests point location in advance, this procedure is affected by the increasing of anonymous users and then apparently a sharp rise in the computational complexity, so its execution time is higher than SDAA. Finally, because of SDAA only needs to calculate the distance between each position, and the distance calculation is transmitted to be the comparison of distances between different users, this comparison improves the computational efficiency, so it has the shortest execution time among these algorithms.

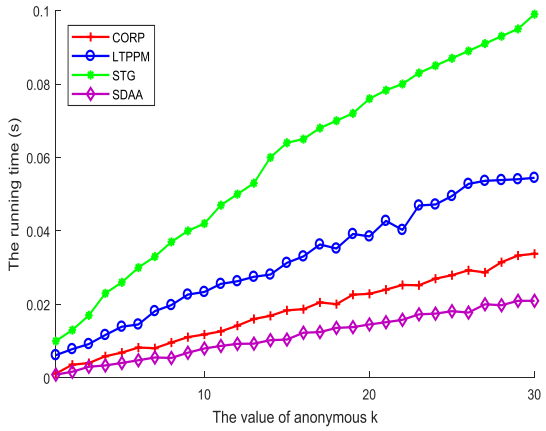

Figure 7 The running time VS. anonymous value *k*

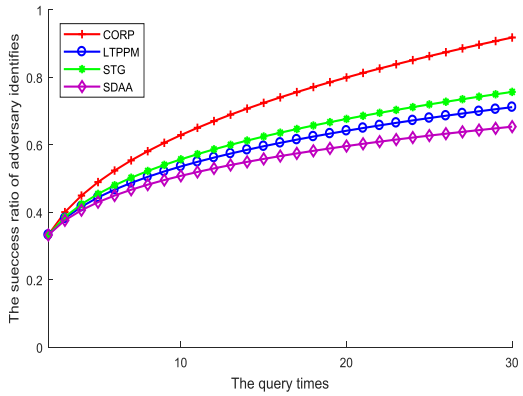

Figure 8 The success ratio of adversary identifies vs. query times

Figure 8 shows the success ratio of the adversary identifies the real user in different number of consecutive queries, and we defined the anonymous value to 20. In general, the more the user queries, the more background knowledge the adversary can acquire, and the higher the adversary's recognition rate to the real user. It can be seen from the figure that with the increasing of the number of queries all algorithms have an increasing ratio to be identified by the adversary. Among these algorithms, CORP algorithm has the highest ratio of identifying the real user, as this algorithm just utilizes precomputation to generalize the real user, but does not effective to deal with the continuous anonymity in successive query about differences between continuous anonymous group, and this leads to the adversary can utilize the differences to guess the real user and obtain the location privacy. Because of the algorithm of STG is mainly designed in Euclidean space this algorithm is vulnerable to be constrained by the road network seriously. Furthermore, as this algorithm utilizes the dummy similar trajectory to generalize the real user and the usability of dummy trajectory is affected by the road network, then within a certain range can't effectively provide privacy protection, so its success ratio of the adversary identifies the real user is higher. LTPPM algorithm utilizes collaboration users that have the same direction to generalize the real user, and this kind of collaborative users are easy to be changed after multiple queries along with the real user, which cause the failure of anonymous and the adversary will have a higher success ratio. Finally, algorithm of SDAA utilizes the central server to calculate the distance to select the same direction, although the conception of this algorithm is similar with LTPPM, the constraint of this algorithm is lower than collaborative users, and more conducive to continuous anonymous. Thus, the success ratio of the adversary identifies the real user is the lowest among these algorithms.

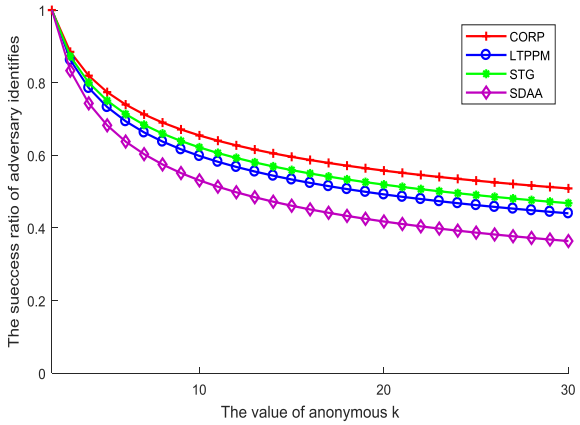

Figure 9 The success ratio of adversary identifies vs. anonymous value *k*

Figure 9 shows the success ratio of the adversary identifies the real user in different anonymous value. In order to facilitate the process of experiment, we confined the value of query time in routing is 10, and the changing process is affected by the increasing of anonymous value. From this figure, we can see success ratios of all algorithms of the adversary identify the real user is descending with the increasing of anonymous value. This is because of that along with the increasing of anonymous value the adversary has to search the real user in a larger region to find the real user and identifies it. So the probability of adversary success identifies descending with the increasing of anonymous value, and the adversary difficult to identify the real user. Among these algorithms, SDAA can select anonymous users with the same direction which reduces the probability of adversary identifying the real user by the correlation of continuous signs along the road, and this process reduces the value of success ratio. Thus, compared with other algorithms, SDAA has the lowest value of success ratio, which means it can provide a better level of privacy protection.

In conclusion, compared with other similar algorithms through above experiments, the algorithm proposed in this paper performs better than them no matter in the capability of privacy protection or the execution efficiency, and it is more suitable to be used in the environment of road network.

6 Conclusion and future works

As the road network is such a special environment that seriously restricts the implementation effect of the privacy protection algorithms designed in Euclidean space, especially when the user in the routing of the continuous query. In order to cope with the problem of

the adversary obtain the privacy by utilizing the differences of anonymous users in each location along the continuous anonymity, and reduce the cost of calculation of the central server. This paper proposes a similar direction anonymous users finding algorithm. With the help of this algorithm, anonymous users with the same direction are selected by the shorted distance, and the algorithm achieves the uniform numbers of anonymous set in each query along the routing, which solve the problem of differences of anonymous users in each location along the continuous anonymity. Furthermore, we utilize the security analysis and experimental validation to further prove that the proposed algorithm is superior in execution efficiency and privacy protection. Of course, this algorithm is not perfect it still limited by the question of whether there are enough anonymous users, so future work will be focused on how to reduce the impact of the number of anonymous users and so on.

Acknowledgment

This work was supported by University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province and Key scientific research projects of Jiamusi in 2017(project number: UNPYSCT-2017149, UNPYSCT-2017175,170034), the National Undergraduate Innovation and Entrepreneurship Training Program of China (201810222033,20180222001).

References

1. Vergara-Laurens I J, Jaimes L G, Labrador M A. Privacy-Preserving Mechanisms for Crowdsensing: Survey and Research Challenges[J]. IEEE Internet of Things Journal, 2016, PP(99):1-1.
2. Shokri R, Theodorakopoulos G, Papadimitratos P, et al. Hiding in the Mobile Crowd: LocationPrivacy through Collaboration[J]. IEEE Transactions on Dependable & Secure Computing, 2014, 11(3):266-279.
3. Chatzikokolakis K, Elsalamouny E, Palamidessi C. Efficient Utility Improvement for Location Privacy[J]. Proceedings on Privacy Enhancing Technologies, 2017, 2017(4) :308-328.
4. Zhang Lei, Ma Chunguang, Yang Songtao, Li Zengpeng. Correlation probability indistinguishable location privacy protection algorithm[J]. Journal on Communications. 2017,(08):37-49.
5. Zhang Lei, Ma Chunguang, Yang Songtao, et al. Location privacy protection model and algorithm based on profiles generalization [J]. Systems engineering and electronics 2016, 38(12): 2894-2900.
6. Palanisamy B, Liu L. Attack-Resilient Mix-zones over Road Networks: Architecture and Algorithms[J]. Mobile Computing IEEE Transactions on, 2015, 14(3):495-508.
7. Zhang Lei, Ma Chun-guang, Yang Song-tao, et al. A real-time similar trajectories generation algorithm for trajectories differences identification resistance [J]. Journal of Harbin Engineering University, 2017, 2017(07):1173-1178.
8. Zhang Lei, Ma Chun-guang, Yang Song-tao, Li Zengpeng. A Privacy Preserving Method from Attacks of Velocity Prediction in Road Network[J]. Journal of Xi'an Jiaotong University. 2017,51(2):27-32.
9. Zeberga K, Jin R, Cho H J, et al. A Safe-Region Approach to a Moving k-RNN Queries in a Directed Road Network[J]. Journal of Circuits Systems & Computers, 2017, 26(05):115-124.
10. Peng T, Liu Q, Meng D C, et al. Collaborative trajectory privacy preserving scheme in location-based services [J]. Information Sciences, 2017, 387(165-179.
11. Zhang L, Li J, Yang S, et al. Privacy Preserving in Cloud Environment for Obstructed Shortest Path Query [J]. Wireless Personal Communications, 2017, 96(2): 2305-2322.
12. Hashem T, Kulik L, Zhang R. Countering overlapping rectangle privacy attack for moving kNN queries [J]. Information Systems, 2013, 38(3): 430-453.
13. Gao S, Ma J F, Shi W S, et al. LTPPM: a location and trajectory privacy protection mechanism in participatory sensing [J]. Wireless Communications & Mobile Computing, 2015, 15(1): 155-169.