Scientific Research Publishing

Journal of Computer and Communications, 2018, 6, 36-47
http://www.scirp.org/journal/jcc
ISSN Online: 2327-5227
ISSN Print: 2327-5219

Trajectory Rotation Privacy Protection Algorithm Based on *k* Anonymity

Zhenpeng Liu[1,2], Xuan Zhao[1], Yawei Dong[3], Bin Zhang[2*]

[1]School of Electronic Information Engineering, Hebei University, Baoding, China
[2]Center for Information Technology, Hebei University, Baoding, China
[3]School of Computer Science and Technology, Hebei University, Baoding, China
Email: *zb@hbu.edu.cn

How to cite this paper: Liu, Z.P., Zhao, X., Dong, Y.W. and Zhang, B. (2018) Trajectory Rotation Privacy Protection Algorithm Based on *k* Anonymity. *Journal of Computer and Communications*, **6**, 36-47.
https://doi.org/10.4236/jcc.2018.62004

Received: January 22, 2018
Accepted: February 10, 2018
Published: February 13, 2018

Abstract

The frequent use of location query services in location-based services will come out a large amount of space-time data related to users. Attackers infer information of location or track based on these rich background knowledge. Therefore, aiming at the problem of trajectory privacy, the context adds instant traffic monitoring based on user behavior patterns, trajectory similarity and other background information. According to the idea of *k* anonymity, proposed a method combined with traffic condition to protect the trajectory privacy. First, the user randomly selects a time point of the real trajectory to rotate to generate dummy trajectory, and then repeat the above process on the real trajectory and dummy trajectory. Up to the generation of $k-1$ dummy trajectory, and according to the actual road conditions and trajectory leakage probability, traversing dummy trajectory to adjust. Finally, it is further proved through experiments that the method will be more efficient and protect privacy well.

Keywords

Location-Based Service, Trajectory Privacy, *k* Anonymity, Traffic Condition

1. Introduction

In recent years, the mobile Internet has brought the development of information technology into a new era, has changed people's traditional way of life, and has affected the medical, entertainment, finance, politics, education and other fields. With the popularity of 4G era, mobile devices such as smart phones have become more portable and real-time, and mobile communication technology is changing with each passing day. Location-Based Service (LBS) is one of the most important service models in the mobile Internet. It integrates positioning tech-

nology, mobile communication technology, internet technology and geographic information system (GIS) technology, and is the most frequently used service used by a mobile internet user. LBS query is divided into two types: snapshot query and continuous query [1]. Snapshot query [2], which can be called passive query, refers to the user issuing query request to LBS service provider and providing the current location. LBS service providers respond to queries and transmit service results. Continuous queries [3] [4] are called active queries. They are users who periodically provide location to service providers to obtain services (such as constantly querying surrounding gas stations, etc.) although location-based services create great convenience for the construction of cities. But sending user's exact location, especially sensitive locations to LBS servers, which puts them at risk of privacy leaks.

At present, researchers put forward many methods for the protection of location privacy, but most of them are to solve the privacy protection of location in snapshot queries, which involves fewer problems in continuous queries. The researchers found that in continuous queries, despite the privacy protection of every position in the trajectory, the attacker can still obtain the approximate trajectory of the user based on the rich background knowledge and the behavior pattern of the user, and thus obtain the approximate trajectory of the user, resulting in leakage of sensitive locations and trajectory. Therefore, the correlation between the location of the locus and its background information becomes an urgent problem to be solved.

The existing methods to protect trajectory privacy can be divided into four categories [5]: generalization method, mixed region method, suppression method, false trajectory perturbation method. Trajectory privacy protection based on generalization method [6] [7] [8] is to generalize an exact location of each moment on an actual trajectory to an anonymous region and send it to an LSP. Generally, the generalized region includes k locations, so that the LSP can not distinguish the authenticity so as to achieve the purpose of trajectory privacy. The method of generalization is mainly to segment the trajectory according to the sampling time, and then to generalize the position. Therefore, the generalization method needs to consider the position correlation attack.

Trajectory privacy protection technology based on hybrid region [9] [10] [11] is mainly used in vehicle networking. Palanisamy *et al.* [11], put forward the method of mix-zone for the first time. The periodic transformation of pseudonyms is used to cut off the correlation between the continuous location and the user, but its privacy protection depends on the number of trajectory in the same anonymous region. Trajectory privacy protection based on suppression method [12] [13] [14] was used in 2004 by suppressing publishing. Grutester *et al.* [12], according to the sensitivity of real location or frequency selectivity. The inhibition method is simple and effective to suppress or delay the updating of the sensitive region. It can handle a part of the trajectory of an attacker, but it can also cause data distortion. Trajectory privacy protection technology based on false trajectory perturbation method [15] [16] [17] [18] is to using TTP or P2P to

realize k anonymity including other users' tracks or historical trajectories and using random way to generate false trajectories to complete k anonymity. Xu *et al.* [15] proposed to use the user history track matching. A method of forming false trajectory. This method is due to the use of historical trajectories. The false trajectory is true and effective, which reduces the identifiability. Wu *et al.* [16] take into account the distance between the true and dummy trajectories and the distance between the dummy and dummy trajectories. At present, the researchers based on the above four privacy protection technologies, more from the user behavior patterns, geographical topology, continuous time position correlation. Position transfer probability and other aspects are considered. In this paper, the user trajectory is rotated to generate $k-1$ dummy trajectory based on dummy trajectory perturbation method. Firstly, the user selects a sampling point randomly from the real trajectory and rotates the point as the center. Then the above process is repeated on the real track and the generating track until the $k-1$ trajectory is generated. If the generated $k-1$ track passes through a more congested section, it can be moved to the nearby road. The leakage probability of traversing the above trajectory and the leakage probability of the real track are shifted again if the difference is large.

2. Related Work

2.1. Request Probability

The probability of request refers to the probability of users sending LBS service requests in a location or a region. The probability of user's service request is predicted by historical trajectory. The map model is divided into a grid of size n × n, and the probability of service request of a grid is:

$$q_i = \frac{\text{number of history service requests in gird } i}{\text{number of history service requests in all girds}} \tag{1}$$

2.2. Leakage Probability

The real trajectory can be represented as $U_0 = \left[(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_m, y_m)\right]$, (x_i, y_i) represents the location of the user at the time of the snapshot query at i time. Trajectory leak probability is the degree of overlap between the location of the request for service at each snapshot query by the user and the position on the corresponding perturbed trajectory. The request probabilities of grids where the ith $(1 \le i \le m)$ position (x_i, y_i) of the real trajectory and its corresponding jth $(1 \le j \le k-1)$ thrashing position (x_{ij}, y_{ij}) are respectively q_i and q_{ij}. When the request probability of these disturbed locations is very close to the request probability of the real location, the attacker can not distinguish the real location. In order to achieve trajectory k anonymity and disrupt the effective position, need to meet the following two conditions:

1) the ith position of the real trajectory does not coincide with the corresponding jth disturbing position, that is, $(x_i, y_i) \ne (x_{ij}, y_{ij})$ for any $j\,(1 \le j \le k-1)$.

2) the difference between the request probability of the ith position of the real trajectory and the corresponding jth disturbance position is less than δ, that is, for any $j\,(1 \le j \le k-1)$, $\left|q_i - q_{ij}\right| = \delta$ is satisfied. δ is user defined, the closer δ is to 0, the closer the probability of request is.

For each snapshot query, the more the effective scrambling location, the smaller the leakage probability. When $k-1$ positions are valid, the leakage probability is the least. The less the effective scrambling location, the greater the leakage probability. When all the $k-1$ positions are invalid, that is, when the k positions are overlapped at one point, the true location information is leaked. Thus, the probability of location leakage is defined as:

$$p_i = \frac{1}{\text{the valid disturbed position of the } i \text{ real position in the snapshot}} \tag{2}$$

The probability of leakage of the trajectory is:

$$P_r = \frac{\sum_{i=1}^{m} p_i}{m} \tag{3}$$

2.3. Trajectory Similarity

Considering the user's behavior habits and the true degree of trajectory, the degree of privacy protection is measured by trajectory similarity, and the higher the similarity between real trajectory and disturbed trajectory, the less discernibility. The real trajectory can be represented as $U_0 = \left[(x_1, y_1), (x_2, y_2), (x_3, y_3), \cdots, (x_m, y_m)\right]$, assuming that the user's position at i time is (x_i, y_i) at an angle θ_i to the initial location (x_1, y_1), so $\tan\theta_i = \frac{y_i - y_1}{x_i - x_1}$. It can be deduced that $\theta_i = \arctan\frac{y_i - y_1}{x_i - x_1}$. Then the real trajectory of the user can be expressed as $U_0 = \left[(x_1, y_1), \langle\theta_1, \theta_2, \theta_3, \cdots, \theta_m\rangle\right]$. Similarly, the generated disturbance trajectory is $T_j = \left[\left(x_{1j}, y_{1j}\right), \left(x_{2j}, y_{2j}\right), \left(x_{3j}, y_{3j}\right), \cdots, \left(x_{mj}, y_{mj}\right)\right]$, which $\theta_{ij} = \arctan\frac{y_i^j - y_1^j}{x_i^j - x_1^j}$. The trajectory similarity is:

$$\sigma^2 = \sum_{j=1}^{k} \frac{E_m\left[\frac{\theta_i^j - \theta_i}{2\pi}\right]}{k} = \frac{\sum_{j=1}^{k}\sum_{i=1}^{m}\left(\frac{\theta_i^j - \theta_i}{2\pi}\right)}{km} \tag{4}$$

When the value range of the σ^2 is [0, 1] and the trajectory similarity is higher, the more similar the disturbing trajectory and the contour of the real trajectory, the worse the resolution.

3. Algorithm to Protect Trajectory Privacy

3.1. System Models

The algorithm based on client-server model is mainly composed of mobile

end-user LBS server. Mobile end-user accesses LBS server through mobile network to obtain map background knowledge and services. The architecture of the system is shown in Figure 1.

The mobile device user is composed of a location/communication module and an intelligent path planning module and trajectory privacy module. The function of the location/communication module is to communicate with the LBS server to obtain the user's position. The path between the user's starting position and the terminating position is the responsibility of the path planning module. To meet the personalized needs of users such as the shortest path avoid congestion avoid the limited area and so on. The work of the trajectory privacy protection module is to rotate the planned path nodes and offset to generate $k-1$ dummy trajectories. Privacy processing of the user's planning path.

3.2. Specific Algorithm

The algorithm proposed in this paper is composed of two main processes: the generation of dummy trajectory and the adjustment of disturbance trajectory. The specific steps are as follows:

1) Random selection of a time point on a real trajectory and rotate θ_i degree around this point, according to the map model, the rotation point is taken as the datum to offset and generate a dummy trajectory T_1'.

2) Choose a random location at all time points of U_0 and T_1', then repeat step 1 to generate the second dummy trajectory.

3) Repeat step 1 ~ 2 until a $k-1$ piece dummy trajectories are generated.

4) Taking into account real-time traffic conditions, traversing all sampling points of dummy trajectories. If the dummy trajectory passes through a congested section, it is shifted to a near non-congested section.

5) Traversing all the sampling points perturbed the trajectory whether they landed in a valid position (the invalid location is rivers, lakes, seas, buildings, etc.). If the sampling point is in the invalid position it will be offset to a valid position. And determine the probability of the request of the corresponding point on the fake track and the real track, if the difference is larger, the point will be offset.

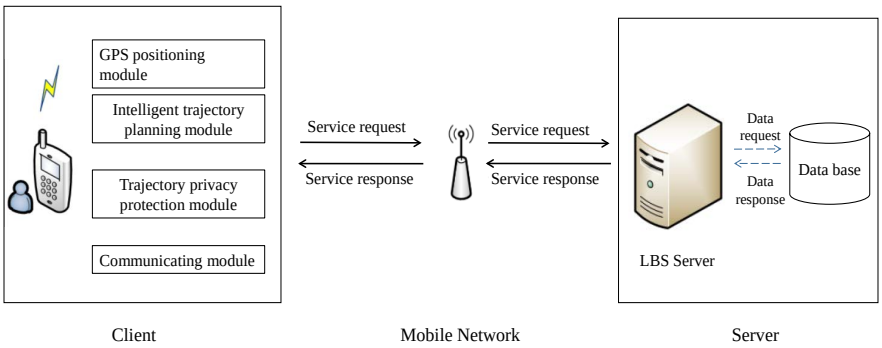


Figure 1. System models.

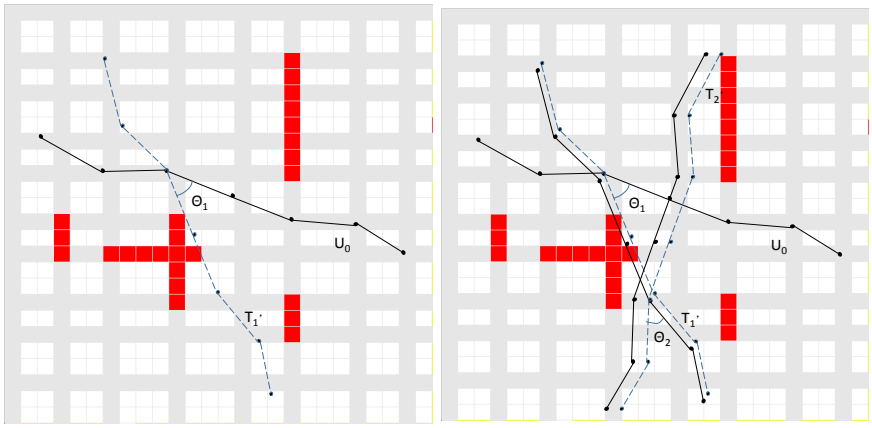
3.2.1. Generate Dummy Trajectories

The process of generating dummy trajectories is shown in Figure 2(a) and Figure 2(b). According to the map background information and real-time traffic on the real trajectory rotation generated $k-1$ dummy trajectory, which realizes trajectory k-anonymous. This paper mainly increases the consideration of real-time traffic, making the generated trajectory more in line with the user's behavior, so that it can not be easily seen by attackers.

In this paper, we set up $k=3$, the gray grid in the diagram represents the valid location, that is, the area where the user can communicate with the LBS service or the area that the user can reach or pass through (it is usually a city road). In the picture, the white grid represents the invalid position, that is, the area where the user cannot reach or pass through, such as lakes, tall buildings, etc. The red grid in the picture represents a section of road where traffic is congested at a certain time.

Step 1: According to the starting point and the end point combined with the real-time traffic conditions to avoid congestion planning the trajectory U_0. According to a certain time interval the trajectory is divided into a number of sampling points. The location of the sampling point can be determined according to the moving speed of the user. The interval can be set to 5 s or 10 s. Then the system sends a query request and receives the result of the query every certain interval. As shown in Figure 2(a).

Step 2: Randomly select a time point in the real trajectory U_0, the corresponding point in time for the center of rotation θ_1 degree. According to the map model, the rotation point is used as the datum to generate the dummy trajectory T_1'. Then, a random position point is selected at all time points corresponding to U_0 and T_1', taking this position as the center of rotation θ_2 degree. According to the map model, the rotation point is used as the basis for migration to generate the second disturbance trajectory T_2', until the $k-1$ disturbance trajectory is generated. The $k=3$ is set up in this paper, so there are two disturbing trajectories. As shown in Figure 2(b).



(a) (b)

Figure 2. Dummy trajectory generation algorithm. (a) Trajectory rotation; (b) Dummy trajectory migration.

3.2.2. Dummy Trajectory Adjustment

Combined with the actual road condition information and map background knowledge, the location points of the generated dummy trajectory are adjusted appropriately to achieve the target of disturbing the attacker. Better protect the user's trajectory privacy. The main process is shown in Figures 3(a)-(c) below.

Step 1: Because of the real trajectory is to avoid congestion, the dummy trajectory should also conform to the behavior habits of the user, according to the actual road information, traversing the $k-1$ dummy trajectory of all sampling points corresponding to the location. If the location point is in the congested section, the location point is shifted to the non-congestion effective point near the nearest request probability (gray grid area). As shown in Figure 3(a).

Step 2: Combining the background information of the map, traversing all the sampling time points corresponding to the $k-1$ dummy trajectory. If its position is in an invalid area, move its location point to the valid location that closest its probabilistic request. As shown in Figure 3(b).

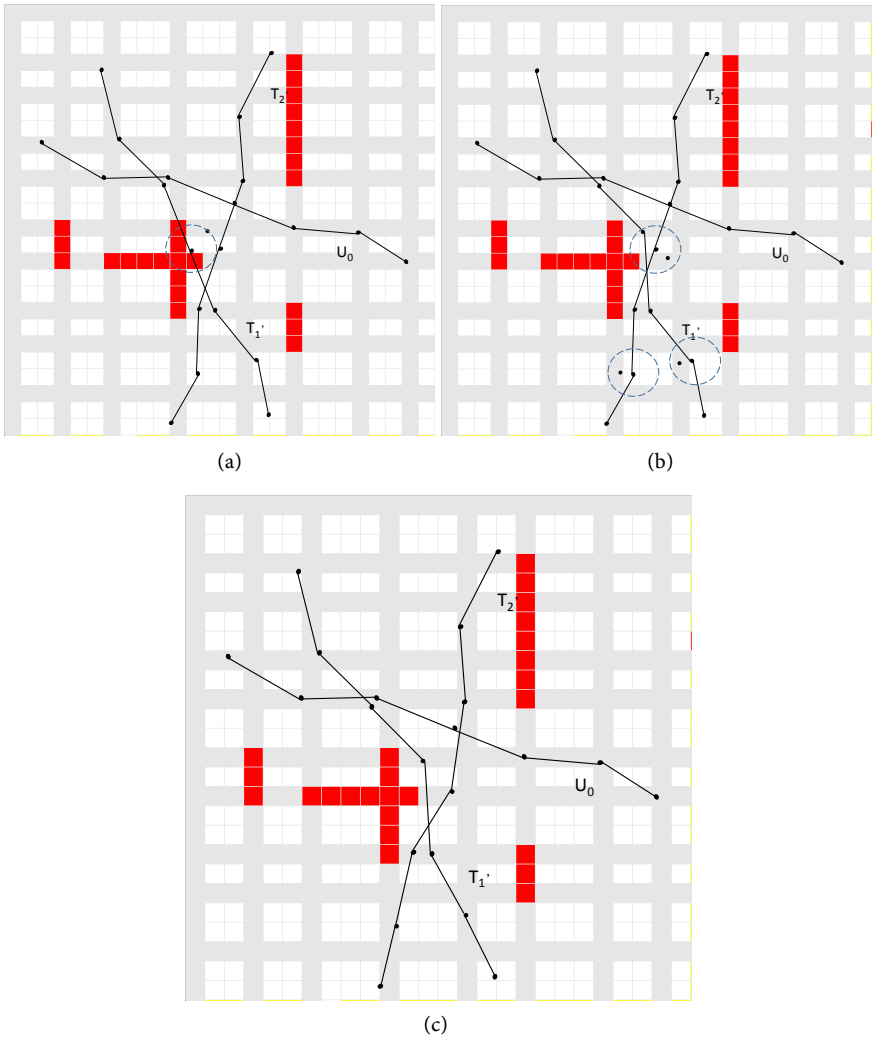


Figure 3. Dummy trajectory adjustment. (a) Trajectory adjustment; (b) Trajectory adjustment; (c) Trajectory adjustment.

Step 3: According to the service request probability of each sampling time point corresponding to the user's real trajectory, the position of all sampling time points corresponding to the $k-1$ dummy trajectory is traversed. If the service request with the real path corresponding to the position of the service request probability is greater than the threshold value of δ, then the center of the service request is the center of that position and the deviation is carried out in the region with radius r. The location based on the map background information should be the position closest to the probability of the true trajectory service request in the virtual circle. As shown in **Figure 3(c)**.

The location of the dummy trajectory is adjusted so that the final disturbance trajectory is closer to the user's behavior habits and can resist all kinds of attacks based on background knowledge.

4. Experiments and Results

The experiment set up the network to complete, in Inter Core i5-7200 CPU 2.7 GHz. In order to make the experimental data sufficient and the experimental results true and reliable, and verify the privacy protection degree and efficiency of this method. The experiment adopts Borlange data set [19] (data set) with high recognition. In this paper, the experiment was carried out on 5000 main roads in the city of Borlange in the range of 6000 m × 6000 m, and the influence of the anonymous parameter k on the anonymity of the trajectories and the efficiency of the algorithm are discussed. In this experiment, the parameter k is set to range from 5 to 30. The parameter m (the number of positions corresponding to the sampling time points on a track) is set to 10. The threshold of request probability difference δ is 0.05. The rotation angle θ of the trajectory is set to $0 < \theta < \pi/4$.

4.1. Experimental Analysis of Privacy Protection Effect

The effectiveness of privacy protection is mainly evaluated by trajectory similarity and leak probability. Trajectory similarity is the contour similarity between dummy trajectory and real trajectory. To a certain extent, it reflects the probability of real trajectory identification. This paper is based on the map knowledge, trajectory similarity and other background information under the premise of adding real-time traffic considerations. In order to test its privacy protection, we compare the similarity of trajectory and the probability of leakage between Li scheme [20] and our algorithm on Borlange data set.

Randomly select 2000 users as the experimental object, repeat the algorithm 100 times. **Figure 4** is a comparative experiment of the effect of k on trajectory similarity. We select the Li algorithm for comparison. The Li method only considers the attacker to master the user's behavior and background information, and based on this, use the rotation method to complete the trajectory k-anonymity. As shown, when the trajectory similarity is 0, it is the optimal value. With the anonymous parameter k increases, the trajectory similarity of

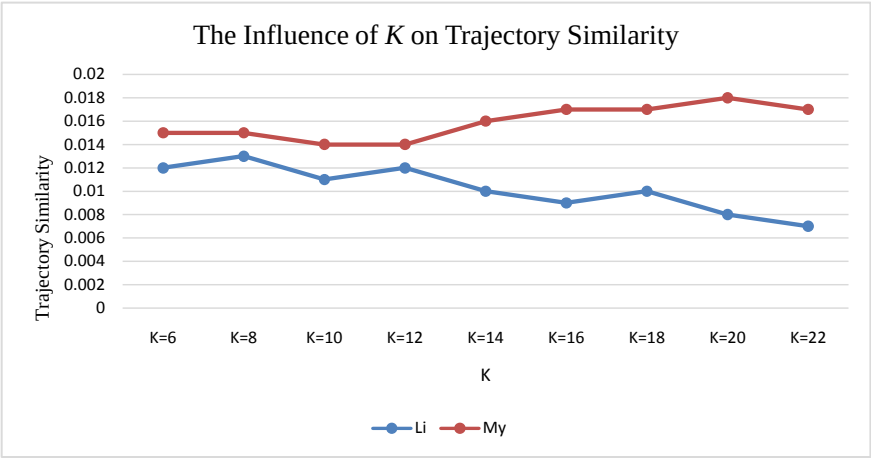


Figure 4. The influence of k on trajectory similarity.

Li algorithm decreases with the increase of k value and gradually becomes stable. The trajectory similarity algorithm proposed in this paper with the k value increasing. Although the trajectory similarity of Li algorithm is less than the trajectory similarity of this paper, but the trajectory similarity of this algorithm is less than 0.02. Because this algorithm adds to the actual road conditions, when dummy trajectory passes through the crowded road, the trajectory is adjusted. So the trajectory similarity will be slightly higher than the Li algorithm. When considering the actual conditions, with the increase of k, the time point at which the dummy trajectory generated needs to be moved increases, and the difference from the contour of the real trajectory increases, but a lot.

Figure 5 is the effect of k on the probability of trajectory leakage. As inferred from 2.3 above, we set $m = 10$, the leakage probability is related to the number of effective perturbed positions s, while s is related to the anonymous parameter k. As shown above, the probability of trajectory leakage between Li and the proposed algorithm increases with k. However, the leakage probability of this paper is slightly smaller than that of Li algorithm, and the leakage probability of this paper is below 20%, which has a good privacy protection effect.

4.2. Operating Time

The operation time of this paper mainly considers two aspects: dummy trajectory generation and dummy trajectory adjustment. The dummy trajectory needs three operations, including the selection, rotation and translation of the time point. Generate a dummy trajectory requires $O(3)$ operations, there for generate $k-1$ dummy trajectory requires time complexity of $O(3k-3)$. The three disturbance trajectory adjustment includes three ergodic processes, and the time complexity is $O(3k-3)$, so the time complexity of the whole algorithm is $O(6k-6)$.

Figure 6 shows the effect of k on the running time. With the increase of k, the running time of the two algorithms increases gradually, but the running time of the algorithm in this paper is a little longer, because the traversal

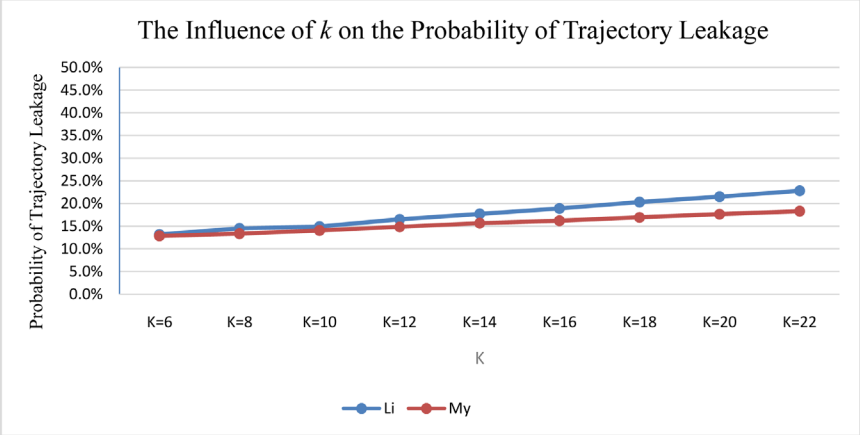


Figure 5. The influence of *k* on the probability of trajectory.

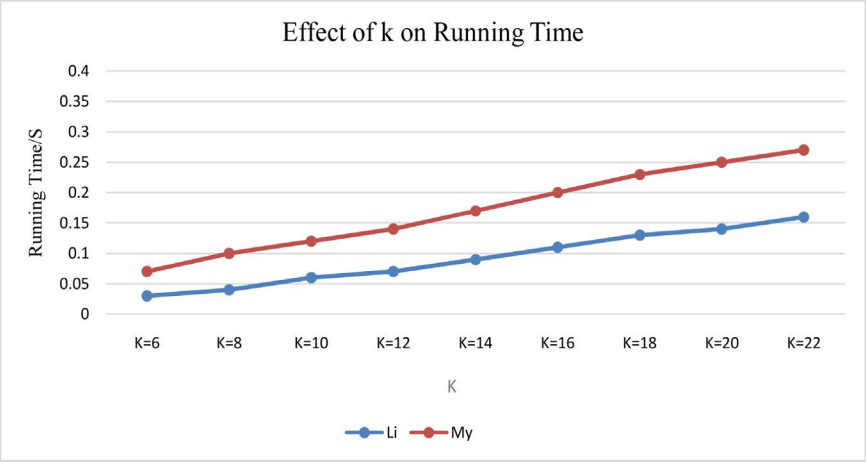


Figure 6. Effect of *k* on running time.

process is added in this paper, which makes the running time of the algorithm increase.

5. Summary

Aiming at the problem of trajectory privacy based on location service. Based on the user behavior patterns and background information, this paper adds the consideration of the actual road conditions and puts forward the real-time traffic rotation *k* anonymous privacy protection method based on trajectory. Combining with the actual conditions, the user first planned trajectory, then $k-1$ trajectories are generated by rotation, and adjusted it according to the actual conditions and the probability of leakage. Finally complete the trajectory *k* anonymity. The performance of this algorithm is tested on the Borlange dataset, which proves that this algorithm has high privacy protection and operating efficiency.

At present, there is no effective way to measure the degree of privacy protection for the trajectory privacy protection in location based service. So the next step is to consider the measurement of privacy protection.

References

[1] Pan, X., Xiao, Z. and Meng, X. (2007) Survey of Location Privacy-Preserving. *Journal of Frontiers of Computer Science and Technology*, **1**, 268-281. (In Chinese)

[2] Zhang, L., Ma, C.G. and Yang, S.T. (2017) Correlation Probability Indistinguishable Location Privacy Protection Algorithm. *Journal on Communications*, **38**, 1631.

[3] Andres, M.E., Bordenabe, N.E., Chatzikokolakis, K. and Palamidessi, C. (2013) Geo-Indistinguishability: Differential Privacy for Location-Based Systems. *Proceedings of the ACM Conference on Computer and Communications Security* (*CCS* 2013), Berlin, 4-8 November 2013, 901-914.

[4] Xiong, W.J., Xu, Z.Q. and Wang, H. (2017) Privacy Level Evaluation of Differential Privacy for Time Series Based on Filtering Theory. *Journal on Communications*, **38**, 1101.

[5] Huo, Z. and Meng, X. (2011) A Survey of Trajectory Privacy Preserving Techniques. *Chinese Journal of Computers*, **34**, 1820-1830. (In Chinese)

[6] Pan, X., Meng, X. and Xu, J. (2009) Distortion-Based Anonymity for Continuous Queries in Location-Based Mobile Services. *Proceedings of the* 17*th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (*GIS* 2009), Seattle, WA, 4-6 November 2009, 256-265.

[7] Bamab, B., Liu, L., Pesti, P. and Wang, T. (2008) Supporting Anonymous Location Queries in Mobile Environments with Privacygrid. *Proceedings of the* 17*th International Conference on World Wide Web* (*WWW* 2008), Beijing, 21-25 April 2008, 237-246. https://doi.org/10.1145/1367497.1367531

[8] Huo, Z., Meng, X. and Huang, Y. (2013) PrivateCheckIn: Trajectory Privacy-Preserving for Check-In Service in MSNS. *Chinese Journal of Computers*, **36**, 716-726. (In Chinese)

[9] Freudiger, J., Raya, M., Felegyhazi, M., Papadimitratos, P. and Hubaux, J.P. (2007) Mix-Zones for Location Privacy in Vehicular Networks. *Proceedings of the First International Workshop on Wireless Network for Intelligent Transportation Systems* (WiN-ITS 2007), Vancouver, August 2007, 1-7.

[10] Krumm, J. (2009) A Survey of Computational Location Privacy. *Personal and Ubiquitous Computing*, **13**, 391-399. https://doi.org/10.1007/s00779-008-0212-5

[11] Palanisamy, B. and Liu, L. (2011) Mobimix: Protecting Location Privacy with Mix-Zones over Road Networks. *Proceedings of the* 27*th International Conference on Data Engineering*, Hannover, 11-16 April 2011, 494-505.

[12] Gruterser, M. and Liu, X. (2004) Protecting Privacy in Continuous Location-Tracking Applications. *IEEE Security and Privacy*, **2**, 28-34. https://doi.org/10.1109/MSECP.2004.1281242

[13] Chen, R., Fung, B.C.M., Mohammed, N., Desai, B.C. and Wang, K. (2013) Privacy-Preserving Trajectory Data Publishing by Local Suppression. *Information Science*, **231**, 83-97. https://doi.org/10.1016/j.ins.2011.07.035

[14] Zhao, J., Zhang, Y., Li, X. and Ma, J. (2014) A Trajectory Privacy Protection Approach via Trajectory Frequency Suppression. *Chinese Journal of Computers*, **37**, 2096-2106. (In Chinese)

[15] Xu, T. and Cai, Y. (2008) Exploring Historical Location Data for Anonymity Preservation in Location-Based Services. *The* 27*th Conference on Computer Communications*, Phoenix, 13-18 April 2008, 547-555.

[16] Wu, X. and Sun, G. (2014) A Novel Dummy-Based Mechanism to Protect Privacy on Trajectories. *IEEE International Conference on Data Mining Workshop*, Shenz-

hen, 14 December 2014, 1120-1125.

[17] Lei, K.Y., Li, X.H. and Liu, H. (2016) Dummy Trajectory Privacy Protection Scheme for Trajectory. *Journal on Communications*, **37**, 2811-2819.

[18] Freudiger, J., Shokri, R. and Hubau, X.J.P. (2012) Evaluating the Privacy Risk of Location-Based Services. Financial Cryptography and Data Security. Springer, Berlin Heidelberg, 31-46. https://doi.org/10.1007/978-3-642-27576-0_3

[19] Nui, B., Li, Q.H., Zhu, X.Y., *et al.* (2015) Enhancing Privacy through Caching in Location-Based Services. *Proceedings of the* 34*th IEEE International Conference on Computer Communications*, Hong Kong, 26 April-1 May 2015, 1017-1025.

[20] Li, F.H., Zhang, C., Nui, B., *et al.* (2015) Efficient Scheme for User's Trajectory Privacy. *Journal on Communications*, **36**, 114-123.