Research on Scheme and Technology for Mobile User’s Trajectory Privacy Protection

Research on Scheme and Technology for Mobile User’s Trajectory Privacy Protection

Ying Tan

Department of Information,Yunnan University of Finance and Economics, Kunming 650221,China
Email: tanyingty@163.com

Abstract. With the development of positioning technologies and wireless communication equipment, the application of location based services is becoming more and more extensive, and a large amount of user mobile trajectory data has been generated. While people enjoy the convenient services brought by these technologies, their own trajectory privacy is also facing great security threats. The article describes and summarizes the system structure, protection scheme, and protection technology in track privacy protection, analyzes and compares it, and points out the future research direction of trajectory privacy protection.

1. Introduction

In recent years, with the rapid development of Internet technology, GPS(Global Positioning System), wireless communication intelligent devices and RFID(Radio Frequency Identification), mobile intelligent terminals are widely used for people to gain convenience in daily life. But when people need to gain these information services, they must provide their location for the LBS(Location-Based Service). If attackers intercepted these location data and string them together, then users’ moving trajectory is formed and accessible. These locations and trajectory imply users’ privacy, including home address, traveling routes, preference in life, health and working conditions. Therefore, privacy protection of location and trajectory data is becoming a new problem and research focus.

2. Relevant concepts

Trajectory refers to a set of location sampling informationTi={idi,(x1,y1,t1),(x2,y2,t2),…,(xn,yn,tn)} set by mobile user Oi over time. In this function,idi is the identity of the mobile users, the location of mobile users at timeti is (xi,yi), and (xi,yi,ti) is a sampling point of the information set Ti.

Trajectory privacy extends the traditional concept of "privacy" to wireless networks, which is composed of sensitive personal information hidden in the trajectory of mobile users.

3. System structure of trajectory privacy protection

3.1 *Central server structure*

A trusted third-party central server is added between the mobile terminal and the LBS server, as shown in Figure 1. First, the user sends the request information to the central server, including user location information, query content, etc. The central server processes the information anonymously through an anonymous algorithm, then sends it to the LBS server and receives the query results returned by the LBS server. Finally, the central server will filter the results back to the users.

Because the central server structure is easy to implement and can effectively reduce the amount of computation on the client side, it has become the most common system architecture mode. The security and anti-attack capability of the central server is the key to this structure. If the central server is broken, the privacy will face enormous risks.

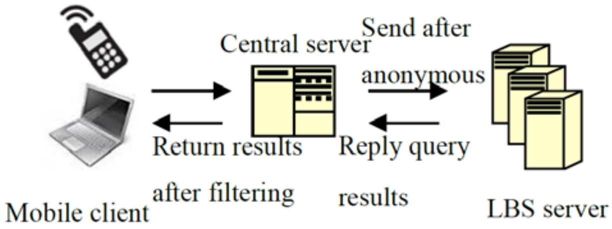

Figure 1: Central server structure

3.2 Point-to-point structure

The point-to-point structure is mainly composed of mobile user terminal and LBS server, in which the responsibility of privacy protection is assumed by the mobile user terminal, as shown in Figure 2. When a user initiates a service request to a LBS server, he will first cooperates with the other users the server collects to complete privacy protection, and then sends the service request to the LBS server. In this structure, the task of privacy protection is mainly completed by the user, so the amount of computation required for the client is larger. In addition, if the credibility between users is not high enough, the risk that privacy faces will also increase.

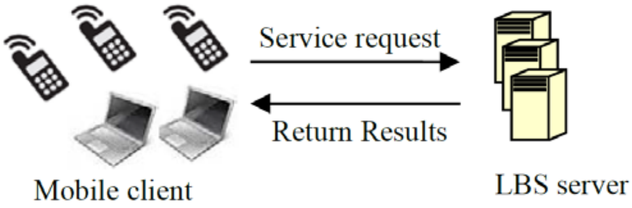

Figure 2: Point-to-point structure

3.3 Comparison of trajectory privacy protection system structure

The comparison of trajectory privacy protection system structure is shown in Table 1.

Table 1: Comparison of trajectory privacy protection system structure

Structures	Advantages	Disadvantages
Central server structure	Easy to implement	Service quality depends on the central server performance
Distributed point-to-point structure	No need for third-party servers	The computation workload of the client is large, and the privacy protection effect is poor.

4. Main solution of trajectory privacy protection

At present, the mainstream solutions can be divided into three categories: privacy protection scheme based on traditional cryptography, scheme based on trusted third party server and distributed scheme based on mobile terminal.

4.1 Encryption scheme

Using cryptography tools to encrypt user's location and query content. However, this encryption process consumes too much resources, when the resources of mobile terminals are very limited, so some experts proposed to put the encryption operations adopted by server and mobile terminals into the cloud. Research on this aspect is still in its infancy.

4.2 Scheme based on trusted third-party server

Mobile users send their service requests to the anonymizer. The anonymizer hides the received information of the users' location and trajectory, and then send it to LBS servers. In this process, the anonymizer must complete the hiding of user trajectory information and the refinement of query results. Therefore, its performance is very important, which will directly affect the privacy protection effect and the service status of users. If anonymous server failures occur, users' privacy will be seriously threatened under this high-dependency protection.

4.3 Distributed scheme based on mobile terminal

The distributed scheme of mobile terminal includes two parts. With the short-range communication technology of mobile terminal, the mobile user makes request, and finds other users who are suitable through the anonymous algorithm. Thus an anonymous area is formed where user actively completing the hiding of his own trajectory. Then select the appropriate user node, forward the anonymous set, and finally feedback the query results to the requesting user. In this process, the attacker obtains a range of areas requested by the user, not the exact location of the user, so as to protect the trajectory privacy of the user. This scheme can effectively avoid the dependency problem in the trusted third-party server scheme, for the users actively participate in the privacy protection of their own trajectory. However, the database and anonymous services are independent from each other, the structure of the architecture is complex, and users need to open Wi-Fi or Bluetooth for a long time, causing relatively large cost of resources, so the active participation of users is not high.

4.4 Comparison of trajectory privacy protection schemes

The comparison of trajectory privacy protection schemes is shown in the Table 2.

Table 2: Comparison of trajectory privacy protection schemes

Schemes	Advantages	Disadvantages
Encryption scheme	Easy to realize	Large resource consumption
Scheme based on trusted third-party server	Good privacy protection effect	Large dependence on the third-party server
Distributed scheme based on mobile terminal	Active protection, independent of third-party servers	Complex structure and large resource consumption

5. Main technical means of trajectory privacy protection

Trajectory privacy is the location privacy of time series, so the most commonly used technology method is to solve the trajectory privacy of mobile users by technology which was used to solve the location privacy protection, and on this basis to improve and develop. The main technical means are: location generalization, false path interference and publication suppression.

5.1 Location generalization

This method extends the user's location and trajectory to fuzzy anonymous area after anonymity processing, so as to protect user's privacy and reduce the probability of identification. This method is mainly through the point-by-point fuzzy processing of sampling points, adding a large number of auxiliary data, so the overhead of the system is large.

The most commonly used and basic technology in generalization is k-anonymity technology. Sweeney et al. first proposed the k-anonymous privacy protection model in 2002, which is mainly used to protect the data in the database. After anonymity, any record in the data table must have at least k-1 records with the same quasi-identifying attribute value, that is, the number of records of any equivalent class in the data table should not be less than k. Then Gruteser used this anonymity technology for relational database to protect location privacy, and proposed k-anonymity location privacy protection model. How to implement anonymous sets, different experts have put forward different implementations, which are mainly divided into three categories: One is to generate k-1 false information and send it to LBS server together with the user's real location or trajectory. For example, JIA et al. proposed to collect k-1 false information through anonymous server. The second is to find other trajectories near the user's real location or trajectory. For example, Nergiz et al. proposed to search for near trajectories in the same equivalence class to form k-anonymous sets and send them to the LBS server together with the real location or trajectory. Thirdly, using the user's own historical real trajectory, such as XU et al.proposed using the user's own historical trajectory as a false trajectory to realize k-anonymous set, thus increasing the difficulty of attackers to find the user's real trajectory.

There are several problems in these schemes. One is about the dependency. The effect of scheme execution depends on the performance of the anonymous server. If the anonymous server fails, users' trajectory privacy will face great risks. Two, when user makes continuous queries, the risk of privacy exposure will increase. As shown in Figure 3, when user A makes continuous query requests at different time points, a continuous anonymous box will be generated. And if an attacker intersects anonymous boxes of different time points, he can infer the user's identification and location.

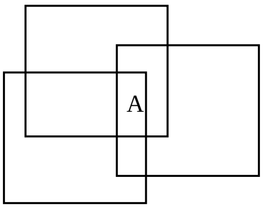

Figure 3: Continuous query anonymous box

To avoid attackers finding real users by intersecting anonymous boxes, Chow et al. proposed that aiming at location privacy protection for continuous queries, User A forms anonymous boxes with other fixed k-1 users at different times, thus avoiding attackers finding user A by intersecting anonymous boxes. However, the problem with this scheme is that when other k-1 users move in opposite directions at a certain time, the anonymous box will be either too large or too small. If it is too large, the system consumption will increase and the service performance will be affected, and if too small, the user location privacy will be exposed.

5.2 False trajectory interference method

By rotating generation method or random production method, false trajectories are generated, and then add these false trajectories or replace true trajectories by false ones. Commonly used technique means of false trajectory interference method includes false location technology and landmark technology. False location technology is to replace the real location of the user by the false, wrong or disguised location, and add it to the service request. The user's sending huge amount of false addresses will lead to increased system consumption. Landmark technology is that users send a landmark location to LBS instead of their real location. The advantage of landmark technology is that attackers can only get the landmark near the location of the user, and cannot get the real and exact location. Therefore, the further the user is from the landmark, the higher the security is. But meanwhile, the far distance from the location will also cause the quality of service requests to drop. Therefore, several problems should be considered in the implementation of false trajectory schemes: 1.The relationship between the generated false trajectory and the real trajectory, such as speed, direction, distance from the real trajectory and so on. These are all the keys to whether the false trajectory can effectively confuse the attacker. 2. How to select the number of false trajectories is also important. If the number is too large, the risk of real users being exposed will be reduced, but the impact on real information will be greater; if too small, the real users will face greater risk. It is an important criterion for the effectiveness of such methods to ensure that the user trajectory privacy is effectively protected and, at the same time, the distortion of real data is controlled within a certain range.

5.3 Publication suppression method

Suppression technology, also known as hiding technology, means suppressingorhiding certain data, publishing the original dataselectively, so that the attacker cannot see the suppressed data. The suppressed data is either deleted or replaced by "*".Therefore, the implementation of suppression technology is also very simple, and the protection of privacy data is very high. Wang Jiabo discussed the use of publication suppression method to protect track privacy. However, the publication suppression method requires distortion processing of data. When the restrained data is too much, the data distortion condition becomes serious, which will greatly reduce the availability of the data. Therefore, how to suppress information and which to suppress are both important for the effectiveness of such kinds of technology. Terrovitis proposed to transform the user's trajectory database so that an attacker cannot infer the user's specific position on the trajectory with a probability higher than a certain one. And if the probability is greater than that level, the position will be suppressed. Gruteser proposed a region-based approach to privacy protection, which divides the area users visit into sensitive and non-sensitive areas. If users enter sensitive areas, their location will be suppressed.

5.4 Comparison of trajectory privacy protection technology

The comparison of trajectory privacy protection technology is shown in the Table 3.

Table 3: comparison of trajectory privacy protection technology

Technical means	Advantages	Disadvantages
Location generalization	The data is real and the quality of service is high.	The overhead of the system is large and the value of k is difficult to determine.
False path interference	Easy to realize and the system consumption is small.	Data distortion, and the number of false trajectories is difficult to determine.
Publication suppression	Easy to realize with high efficiency	Data distortion and availability reduction

5.5 Personalized trajectory privacy protection

The trajectory contains a lot of information. When the user initiates the query service, some sensitive information is available for some trajectory applications but not for other ones. So, trajectory privacy is also a very personalized problem. Aris et al. discussed personalized privacy and considered personalization in their proposed k-anonymity model of user trajectory in LBS. XiongShengchaoet al. proposed that users can choose the privacy-sensitive trajectory fragments independently, and authorize access to different trajectories, so that the location sampling points in the invisible privacy-sensitive fragments can be reasonably dispersed into multiple locations which are frequently accessed around. Li Wenping et al. proposed that the sensitive trajectory and the insensitive trajectory together should generate the hidden variable first, and then we can generate the random trajectory according to the hidden variable to replace the sensitive trajectory. In fact, the protection of privacy for personalized trajectory is essentially publishing trajectories that users deem insensitive and protecting ones that users believe sensitive. Then the relationship between sensitive trajectory and insensitive trajectory, the hiding of sensitive information, the influence of fuzzy or generalization on the quality of trajectory service, and the consumption of system are all the issues that must be considered in personalized trajectory privacy protection. In 2007, Machanavajjha et al. proposed an l-diversity model based on k-anonymity, CHOW et al. proposed that the l-diversity idea can be applied to personalized privacy protection in trajectory privacy protection. That is, the k-trajectories in an anonymous set must be diverse, and the difference between trajectories is used to ensure the security of privacy. But how to determine the difference of trajectory space and how to prevent the privacy leakage caused by the lack of difference will be a serious problem.

6. Closing remarks

The more precise the location people provide to the location server, the better the service they get, but at the same time, the greater the threat to the location and trajectory of the user. How to make a balance of which users want to obtain the best service and also want to maximize the protection of privacy. This is the biggest problem we are facing. In addition, the research on the personalization of trajectory privacy service is still in its infancy, which is also a key direction for further research.

7. Reference

[1]Riboni D, Pareschi L, Bettini C. Shadow attacks on user' anonymity in pervasive computing environments[J]. Pervasive and Mobile Computing,2008,4(6):819-835

[2]Sun X, Sun L, Wang H. Extended k-anonymity models against sensitive attribute disclosure [J]. Computer Communications , 2011,34(4):526-535

[3]Hu G W, Yang J. Research Progress of Track Privacy Protection Technology [J]. Computer science.2016, 43(4):16-23

[4]Tramp S, Frischmuth P, Arndt N. Weaving a Distributed, Semantic Social Network for Mobile Users[C]//8th Extended Semantic Web Conference, 2011:200-214

[5]Sweeney L. Achieving K-anonymity privacy protection using generalization and suppression[J]. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 2002,10(5):571-588

[6]Gruteser M, Grunwald D. Anonymous usage of location-based services through spatial and temporal cloaking[C]//Proceedings of the International Conference on Mobile System, pplications and Services. San Franciscos,USA,2003:163-168

[7]JIA J,ZHANGF.Nonexposure accurate location k-anonymity algorithm in LBS[J],The Scientific Word Journal,2014,2014(1):619357

[8]NERGIZ M E,ATZORI M,SAYGIN Y,etal.Towards trajectory anonymization:a generalization-based approach[J].Transactions on Data Privacy,2009,2(1):47-55.

[9]XU T,CAIY.Exploring historical location data for anonymity preservation in location-based services[A].Proceedings of the 27th IEEE International Conference on Computer Communications 2008[C].Phoenix,USA,2008.1220-1228.

[10]William E Winkler. Using simulated annealing for k-anonymity [R]. Research Report 2002-07, US Census Bureau Statistical Research, Division, 2002

[11] Song J L, Liu G H, Huang L M. Optimization algorithm of K value in k- anonymous privacy protection model [J]. Minicomputer system.2011，10: 1987-1993

[12]Chow C Y,Mokbel M F,Enabling privacy continuous queries for revealed user locations[C]//Anonymous Proceedings of the 10th International Symposium on Advances in Spatial and Temporal Database.Boston:Springer,2007:258-275.

[13]LU R,LIN X,LIANGX, et al. A dynamic privacy-preserving key management scheme for location-based services in vanets[J].IEEE Transactions on Intelligent Transportation System,2012,13(1):127-139.

[14]SHIN K G,JU X,CHEN Z, et al.Privacy protection for users of location-based services[J].IEEE Wireless Communications, 2012.19(1):30-39.

[15] Wang J B. Research on location-based service trajectory privacy protection technology [D].HangZhou: Electronic University Of Science &Technology Of Hangzhou，2014.

[16]TerrovitisM,Mamoulis N. Privacy preserving in the publication of trajectories//Proceedings of the 9th International Conference on Mobile Data Management(MDM 2008).Beijing,2008:65-72.

[17]GruteserM,Liu X. Protecting privacy in continuous locational-tracking applications. IEEE Security and Privacy,2004,2(2):28-34.

[18]Aris G D,Vassilios S V,Panayiotis B.A network aware privacy model for online requests in trajectory data[J].Data & Knowledge Engineering,2009,68(11):431-452.

[19] Xiong S C, Wu X, Peng Z Y. A fine-grained trajectory privacy preserving scheme to maintain data availability [J]. Journal of East China Normal University (Natural Science Edition)，2015（5）: 96-103.

[20] Li W P, Yang J , Zhang J P. Personalized trajectory privacy protection algorithm based on CCA [J]. Journal of Jilin University (Engineering Edition).2015（2）: 630-638.

[21]Machanavajjhala A，Gehrke J, Kifer D. L-Diversity: Privacy beyond k-anonymity[C]. Pro. Of the 22nd International Conference on Data Engineering Atlanta, Georgia, USA:IEEE Press, 2006:24-35.

[22]CHOW C Y,MOKBE M F,LIU X. Spatial cloaking for anonymous location-based services in mobile peer-to-peer nvironments[J].GroInformatica,2011,15(2):351-380.