Users' Privacy Protection Scheme in Location Based Services

Tu-Liang Lin
Department of Management Information Systems,
National Chiayi University
No. 580 Sinmin Rd.
Chiayi City, Taiwan
Tel:886-5-2732906
tuliang@mail.ncyu.edu.tw

Pin-Jie Wang
Department of Management Information Systems,
National Chiayi University
No. 580 Sinmin Rd.
Chiayi City, Taiwan
Tel:886-5-2732906
wes830129@gmail.com

ABSTRACT
LBS (Location Based Service) has become more and more popular in recent years. LBS allows Users to explore the surrounding environment. Users need to transmit their information such as position, POI, identification to untrusted service provider who can be considered as an attacker. By using those data, the attacker may obtain users' privacy. Therefore, there is an increasing amount of research to protect the privacy of users when using the LBSs. Our research aims to improve the shortcomings of past protection methods. In the experiment, our approach is compared with other approaches by using entropy and our approach can provide better protection in three kinds of side information applications.

CCS Concepts
• Security and privacy → Software and application security → Social network security and privacy

Keywords
Location Based Service; Location Privacy; Dummy Based Protection.

1. INTRODUCTION
When users using the location based service, their information will be collected by the untrusted service provider which can be considered as an attacker. They might illegally abuse users' information, such as leaking users' personal data to third parties, causing users privacy been invaded. In this research, we choose mobile device-based architecture and adopting dummy location techniques to design our approach. The dummy location based technique protects location privacy by selecting other locations on the map to assist the users to hide their positions. According to the experimental results, our approach can select better candidates, so that, the real position cannot be identified by the attackers so easily.

ICEMT 2018, July 2–4, 2018, Okinawa, Japan

ACM ISBN 978-1-4503-6525-3/18/07…$15.00
https://doi.org/10.1145/3206129.3239418

2. RELATED WORK

2.1 The Goal of Protection
There are two types of protections in LBS privacy, (1) query privacy protection and (2) location privacy protection [1]. Query privacy refers to the relationship between the user's own private information and the contents of the LBS requests. For example, when a user's queries are always related to alcohol, the attacker might infer that this user has a drinking problem, and this user might develop a disease caused by drinking too much alcohol in the future. Location privacy refers to the relationship between the user's own private information and the location information in LBS requests. For example, a user always send LBS requests near a bar, the attacker might also infer that this user has a drinking habit.

2.2 The Protection Architecture
Depending on how the protection mechanism is built, we can divide the protection architectures into two categories: TTP (Trusted Third Party) and mobile device-based. In TTP, (see Figure 1), there are third parties which is trusted by all users and can provide protection. If a user wants to send LBS requests, he needs to send the request to the third party, then the third party will transfer the requests to the service provider. The service provider will response the requests to the third party, then third party will offer corresponding information to the user. The drawback of TTP architecture is single point of failure. Because the attacker realizes that every user will send requests to the TTP first, TTP will be considered as the target of the attack. In addition, as all requests will be sent to the TTP for protection, the TTP system is overloaded. Once the TTP collapsed, the protection mechanism will be no longer existed. The risk of user privacy being violated increases. On the other hand, in mobile device-based, (see Figure 2), the protection mechanism depends on user's own mobile phone, so it can prevent the problems in the TTP architecture.

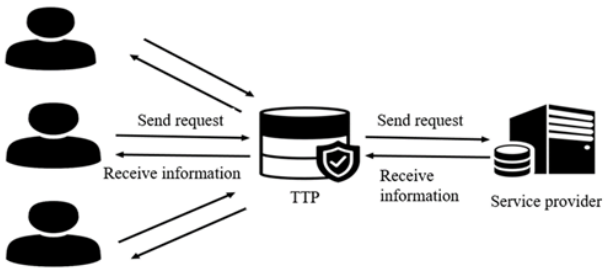

Figure 1. Architecture of Trusted Third Party (TTP)

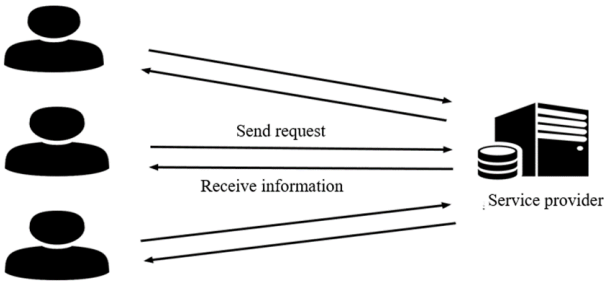

Figure 2. Mobile device-based architecture

2.3 Protection Techniques

In order to deal with the privacy problem, many approaches have been proposed. A policy-based approach is to constrain location service providers (Google, Baidu, etc.) through the development of privacy protection practices or standards that can limit the service providers to disclose users' information or to prevent improper usage. Cryptography primitive-based approaches [2] mainly use the cryptography-related technology to encrypt the information between users and the service provider, so that although the attacker can obtain the information, they still can't easily decrypt the true information. Location perturbation and obfuscation techniques [3] reduce position accuracy by adding noises. It makes the real position changes within a certain range or replaces the real position with a fake position, then sending the adjusted position to service provider. Spatial and temporal cloaking techniques [4] will form a virtual cloaking region which contains the users who need protections and it will send one of the position in the cloaking region instead of the real position. Dummy location techniques [5] produce a variety of different dummy locations, then sending dummy locations and real location together to the service provider, so it will make the attacker difficult to distinguish the real one.

2.4 Previous Research

2.4.1 DLS

Our research belongs to dummy location techniques. In 2014, Niu et al. proposed DLS [5], which can protect user's location privacy by selecting some dummy locations. Unlike previous studies, DLS considered the side information when selecting dummy locations. Side information means the information which can be used to remove some unsuitable dummy locations. In their research, they use the probability of sending LBS request in each location as side information. We will demonstrate how the attacker uses side information to remove unsuitable dummy locations. (see Figure 3), different styles of the grid represent different request probability. The marked grid will be used as dummy locations. We compare marked grid (1) and grid (2), the probability of request in grid (2) is much more than grid (1), which means in grid (1) is Impossible to send the request, so that, the attacker might remove grid (1). There would have been five dummy locations to provide protection originally, but the attacker can remove some unsuitable dummy locations by using the side information, so that protection effect will decrease. DLS (see Figure 4) will select the dummy locations which probability is the same or familiar with real location.

2.4.2 TTcloak

In 2015, Niu et al. proposed TTcloak[6], which also protects user's location privacy by selecting dummy locations. The different between DLS and TTcloak is that TTcloak uses different type of side information and TTcloak also protect query privacy. In TTcloak, Niu et al. use request type as side information. The demonstration is shown below (see Figure 5). Different icons in the grid indicate different request types, and different numbers of icons indicate the frequency of past requests of the corresponding type. Suppose the user request geographic information request for train type, user will remove the grids which probability are not the same with real one, then user compare the different type of request in real grid in order to choose suitable query type to protect real type. Suppose the user request geographic information request for train type, user will remove the grids which probability are not the same with real one, then user compare the different type of request in real grid in order to choose suitable query type to protect real type. After comparing, user will choose restaurant type as a removal factor, then removing the grids which probability are not similar with real one. Finally, the grids in Fig. 5 without shade will be the candidates of dummy location.

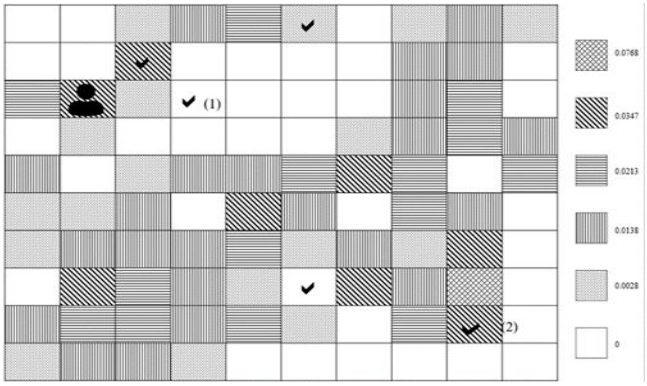

Figure 3. Choosing dummy location without side information

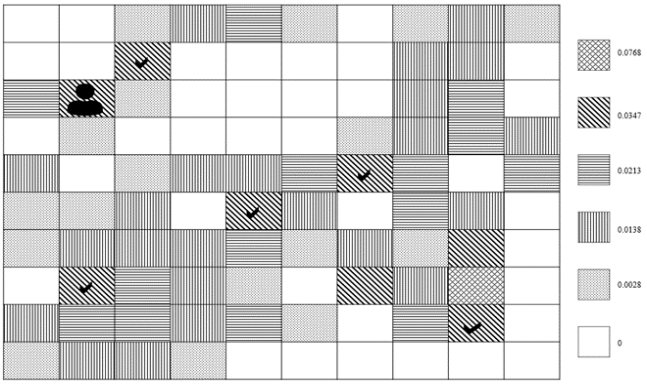

Figure 4. DLS, Choosing dummy location with the side information

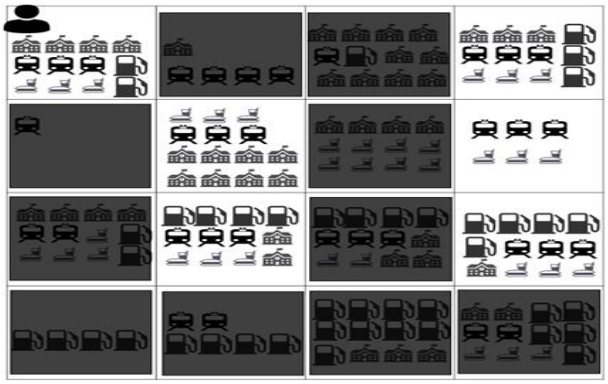

Figure 5. TTcloak approach

3. OUR PROPOSED METHOD

3.1 Our Method

According to past Niu's et al. researches, the side information obtained by using the times of sending request in single grid divide by the times of sending request in all grid. They don't consider the situation of sending request in single grid, and it might cause the following situation. User sends request of train station in grid (5) (see Figure 6), TTcloak will choose grid (1), (3), (7), (8) as candidates of dummy location. In grid (1) and (7), the attacker will find that the probabilities of sending train station and restaurant is much less than other type of request in corresponding grid, so the attacker will remove this kind of grid and it will cause the protection effect decreased.

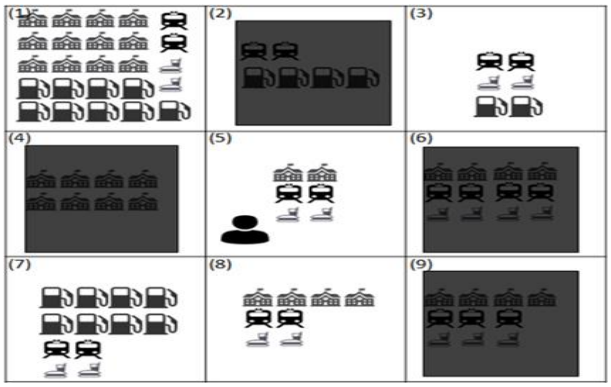

Figure 6. The situation of request distribution

In order to prevent this kind of attack, we will consider situation of sending probability in each single grid. For example, user will send a request to location service provider to obtain train station information (see Figure 7). First, we will pick the grids which have the same total probability of sending with the real position as candidate set, as following (see Figure 8), the grid without shadow will be chosen as candidates. Second, we use the probability of single type of request as condition to pick up the suitable grids from previous candidates. (see Figure 9), we use the probability of sending request about train station as condition to remove unsuitable candidates. The remaining grid without shadow will be chosen as candidates. In order to protect query privacy, we will choose a kind of query type which sending probability is the same with the real one. Then sending it together with real type in order to confuse attacker. So we will use the probability of fake one as condition to remove unsuitable from previous candidates, the remaining grid without shadow (see Figure 10) will form the final candidates.

3.2 Location Privacy Metric

How to measure a protection method is a very important issue in LBS privacy protection. There are two types of measurement methods, respectively uncertainty-based approach [7] and distortion-based approach [8]. The distortion-based approach which uses the estimation error between the real location of the user and the location which is inferred by the attacker as a metric. For example, the real user's coordinate is (x_r, y_r). The attacker will infer a coordinate (x_f, y_f) by using the information which sent by user. We can use Euclidean distance to calculate the distance difference. If the result of calculate is close to zero, it means that the protection mechanism can't protect user effectively. The uncertainty-based approach means ability of determining the real location from a set of candidates by the attacker. Recently, entropy is considered a good measurement for uncertainty. We will use the throwing coin incident to explain entropy. Suppose there are two coins, one of which is uniform, the probability of occurrence of positive and negative is equal, the other is inconsistent, and the positive probability is 70% and the negative probability is 30%. It hard for guessing the result of the first coin because the incident probability is the same, but for the second coin, it is easy to guess the result. Entropy use the incident probability to obtain a value. The larger the value represents the higher the uncertainty.

The entropy formula is presented below.

$$\mathrm{H} = -\sum_{i=1}^{n} P(x_i) * \log_2 P(x_i) \tag{1}$$

H is entropy, n is the number of events, P (X_i) is the probability of occurrence of an event, i represents the i-th event.

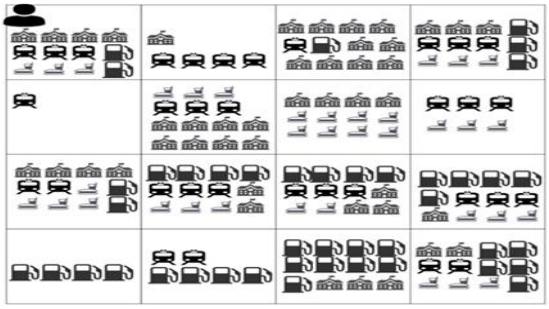

Figure 7. The situation of request distribution

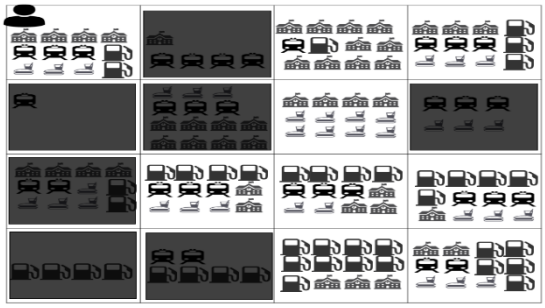

Figure 8. Choosing dummy location by using total probability of sending request

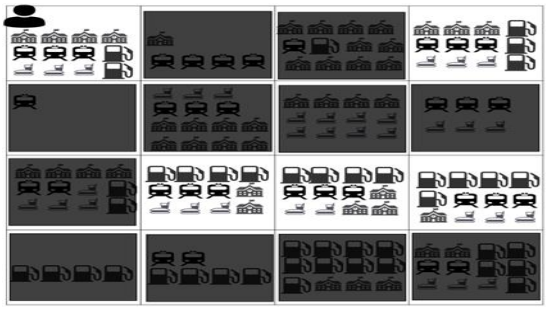

Figure 9. Choosing dummy location by using total probability of single type of sending request

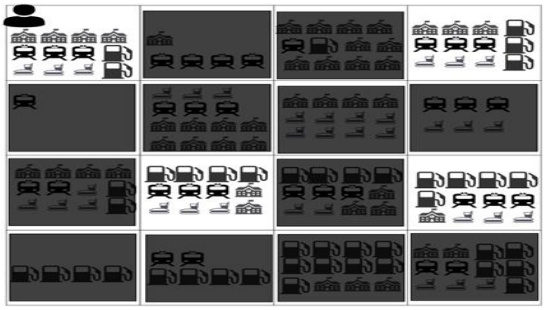

Figure 10. Choosing dummy location by using total probability of dummy type of sending request

4. PERFORMANCE EVALUATIONS

4.1 Simulation Setup

In our research, we also use k-anonymity [9] and l-diversity [10] to determine the anonymity set. K-anonymity means the real location can hide in k-1 dummy location. L-diversity means that there will be l-1 type of request to protect real one.

In the simulation, we divide the map into 100 grids and there are six kinds of request. We assign different frequency for different types of requests in each grid. There are two parameter used in our simulation. K is related to k-anonymity, and is set from 2 to 10. L is related to l-diversity, and is set 2 which means we will choose one kind of request to protect real one. There are three types of side information applications in our experiment to choose the dummy location candidates. First, we use the total probability of sending request in each grid as side information. The second one is using request type as side information. The final one is using the situation of sending request in single grid as side information. Our method was compared with DLS, TTcloak, Random, Optimal by using entropy. The higher entropy means the method can provide better protection. In optimal approach has the highest value which generated by sending k location with the same probability.

Before the entropy calculation, we need to convert the probability values through the normalized transformation, so that these probability values can be used to evaluate the merits of the selected virtual location candidate sets. The normalized formula is presented below.

$$P_i = \frac{q_i}{\sum_{l=1}^{k} q_l} \quad , i = 1,2, \dots, k \tag{2}$$

q_i means the probability that the i-th grid in the candidate set sends the request. p_i is the result of the i-th grid after normalizing and the total of p_i is 1. Then, we use the normalized result to calculated entropy.

4.2 Experiment Result

In first side information application (see Figure 11), it shows that as k increases, the larger the entropy value. Our approach can provide better protection than other approaches in this side information application. But the random approach which determines dummy location set by randomly choosing is quite close DLS, TTclock, our approach. Because our data is randomly generated, it may not present significant difference with other approaches. In second side information application (see Figure 12), our approach is significantly different from and TTclock and DLS, especially after k = 6. Because DLS didn't consider the probability of sending type, it can't not provide enough protection when the attacker using this kind of side information to infer the unsuitable dummy locations. In final side information application, (see Figure 13), it shows that our approach also can provide better performance in the application, because the approaches in past didn't notice the situation of sending request in each grid.

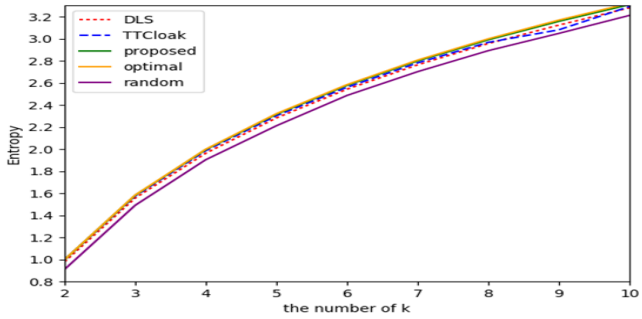

Figure 11. Result of first side information application

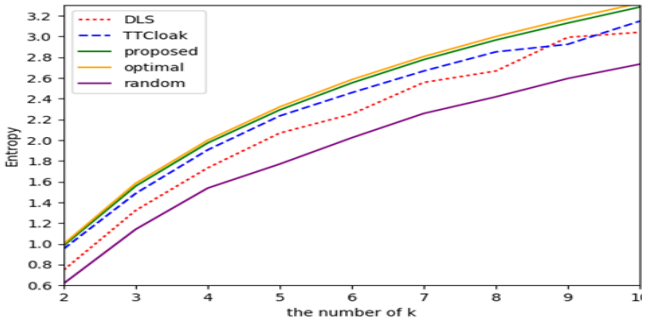

Figure 12. Result of second side information application

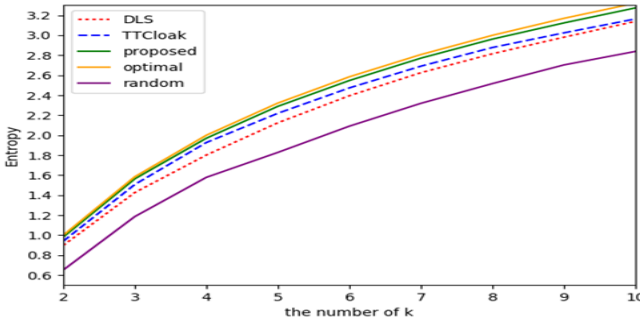

Figure 13. Result of final side information application

5. CONCLUSIONS

In this paper, we proposed a problem when using side information to choose dummy locations. The attacker can use this kind of side

information to remove some dummy locations which are used to protect user's location privacy. According to the experiment result, our approach can provide better entropy metric which means the attacker can't recognize the unsuitable dummy locations easily. In other words, users can enjoy LBS, and suffer less privacy violations

6. ACKNOWLEDGMENTS

This paper is supported by the Ministry of Science and Technology (MOST) of Taiwan. MOST provides the research funding and devices. The related project number of this work is MOST- 106-2221-E-415 -001 -.

7. REFERENCES

[1] Shin, K. G., Ju, X., Chen, Z., and Hu, X. 2012. Privacy protection for users of location-based services. Wireless Communications. 19, 1 (Feb. 2012), 30-39. IEEE. DOI= https://doi.org/10.1109/MWC.2012.6155874

[2] Bilogrevic, I., Jadliwala, M., Kalkan, K., Hubaux, J. P., and Aad, I. 2011. Privacy in mobile computing for location-sharing-based services. International Symposium on Privacy Enhancing Technologies Symposium 77-96. Springer, Berlin, Heidelberg. DOI=https://doi.org/10.1007/978-3-642-22263-4_5

[3] Peddinti, S. T., Dsouza, A., and Saxena, N. 2011. Cover locations: availing location-based services without revealing the location. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society* (Chicago, Illinois, USA, Oct. 17-17, 2011). WPES '11. ACM, New York, NY, 143-152. DOI=http://doi.acm.org/10.1145/2046556.2046576

[4] Lu, H., Jensen, C. S., and Yiu, M. L. 2008. Pad: privacy-area aware, dummy-based location privacy in mobile services. In *Proceedings of the Seventh ACM International Workshop on Data Engineering for Wireless and Mobile Access* (Vancouver, Canada, June 13 - 13, 2008). MobiDE '08. ACM, New York, NY, 16-23. DOI= http://doi.acm.org/10.1145/1626536.1626540

[5] Niu, B., Li, Q., Zhu, X., Cao, G., and Li, H. 2014. Achieving k-anonymity in privacy-aware location-based services. In *Proceedings of the INFOCOM* (Toronto, ON, Canada, 27 April-2 May, 2014). IEEE, 754-762. DOI= https://doi.org/10.1109/INFOCOM.2014.6848002

[6] Niu, B., Zhu, X., Li, W., Li, H., Wang, Y., and Lu, Z. 2015.A personalized two-tier cloaking scheme for privacy-aware location-based services. *In Proceedings of the Computing, Networking and Communications* (Garden Grove, CA, USA, Feb. 16-19, 2015), IEEE, 94-98. DOI= https://doi.org/10.1109/ICCNC.2015.7069322

[7] Gruteser, M., & Grunwald, D. 2003. Anonymous usage of location-based services through spatial and temporal cloaking. I*n Proceedings of the 1st international conference on Mobile systems, applications and services* (San Francisco, California, May 05 - 08, 2003). MobiSys '03. ACM, New York, NY, 31-42. DOI= https://doi.org/10.1145/1066116.1189037

[8] Shokri, R., Theodorakopoulos, G., Le Boudec, J. Y., and Hubaux, J. P. 2011. Quantifying location privacy. In *Proceedings of the Security and privacy* (Berkeley, CA, USA, May 22-25, 2011), IEEE, 247-262. DOI= https://doi.org/10.1109/SP.2011.18

[9] Sweeney, L. 2002. k-anonymity: A model for protecting privacy. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 10, 05(October 2002), 557-570. DOI= https://doi.org/10.1142/S0218488502001648

[10] Machanavajjhala, A., Gehrke, J., Kifer, D., and Venkitasubramaniam, M. 2006. l-diversity: Privacy beyond k-anonymity. *In Proceedings of the 22nd International Conference on Data Engineering* (Atlanta, GA, USA, USA, April 3-7, 2006). IEEE, 24-24. DOI= https://doi.org/10.1109/ICDE.2006.1