Journal of Electronics and Information Science (2024)
Clausius Scientific Press, Canada

DOI: 10.23977/jeis.2024.090303
ISSN 2371-9524 Vol. 9 Num. 3

A Method for Generating Dummy Location Based on Spatiotemporal Correlation

Tingting Gao, Jiaxiang Gao, Yiwei Liao, Shenglin Wang

Institute of Computer Science and Information Engineering, Harbin Normal University, Harbin, 150025, China
gaotingting@stu.hrbnu.edu.cn

Keywords: Location Privacy; Privacy Protection Methods; Spatiotemporal Correlation; Dummy Location

Abstract: To improve the concealment of the dummy location generated by the dummy location generation algorithm in location privacy and enhance the ability to resist the attacker to deduce the user's real location based on background knowledge, this paper proposes a dummy location generation method based on spatiotemporal correlation. Firstly, the historical query probability is used to preliminarily screen and select the appropriate dummy location set, and then the spurious location set is filtered and optimized by adding the relevant definition of spatial sensitivity metric, and the dummy location with higher concealment is filtered and generated. Then, through theoretical analysis and demonstration, it is verified that the algorithm can effectively resist the attacker's background knowledge inference attack, to improve the concealment of the dummy location. Finally, the future research directions of the dummy location generation method are summarized and prospected.

1. Introduction

With the rapid developments of positioning capabilities and the widespread use of wireless networks technology, the location-based service (LBS) has come into our daily life [1]. More and more location-based applications have emerged providing various services for people's work and daily life needs. For example, visitors can send Point of Interest query to the LBS servers. Game players can share their game positions and scores with others nearby. However, despite the great convenience supplied by LBS, it introduces serious challenges of personal privacy. When a user sends a query to the LBS servers, the untrusted servers may collect the users' personal details surreptitiously including location information and queried interests. Then the untrusted servers can track the user or release the user's personal information to others, which may cause potential damage to the user. Thus, we need to take appropriate measures to protect users' location privacy.

In the past few years, the problem of location privacy leakage in LBS has received extensive attention from scholars at home and abroad, and some solutions have been proposed. For example, K-anonymization method, dummy location method, coordinate transformation method [2], cryptography method, differential privacy method and so on. Among the types of location privacy protection methods, the dummy location-based method has the advantages of low computational cost,

does not rely on third parties, and allows users to obtain accurate query results, so it is one of the most frequently used LBS location privacy protection methods [3].

The basic idea of this method is to allow users to use LBS services under the premise of protecting their location privacy. Specifically, the user submits a location collection containing the real location and a series of dummy locations to the LBS service provider, which hides and protects the user's real location [4]. Meanwhile, the user can still use the LBS service through this location collection. Therefore, in this paper, by analyzing the location privacy leakage problem faced by users using LBS and the existing location privacy protection techniques of LBS, we systematically study the dummy location method in the user location privacy protection techniques in LBS and propose a dummy location generation method based on spatiotemporal correlation. The algorithm effectively protects the location privacy of users when using LBS, meets the practical needs of users using LBS in their daily life, and improves the satisfaction of users' experience of using LBS. It also actively promotes the progress and development of LBS and its application services and has a far-reaching impact on the realization of intelligent services.

2. Related work

2.1 Research status

The main idea of constructing dummy locations is to send the user's real location to the LSP together with many dummy locations forming an anonymised set. This approach does not require a third-party anonymised server to join in and avoids the risk of privacy leakage due to the trust issues of anonymised servers or attacks on anonymised servers. However, since it does not consider factors such as query probability, it is vulnerable to attacks on side information with background knowledge such as the user's s historical query probability, map information, etc., there is a serious risk of privacy leakage.

To this end, reference [5] achieves trajectory blurring by forming k-1 sensitive regions (point of interesting, POI) similar to the sensitive points, combining motion patterns, road network topology, and road weights within the sensitive regions, however, the privacy preserving effect is difficult to be guaranteed for adversaries with background knowledge. reference [6] proposes a k-anonymous location privacy protection scheme based on virtual location and Stackelberg game, which achieves mutual optimisation of user-adversary goals and resists single-point attacks and inference attacks by constructing a dummy location set using the Stackelberg game framework, but a semi-trustworthy third party is prone to cause a bottleneck in system performance. To prevent attackers from using query information to infer and carry out attacks, reference [7] uses Voronoi diagrams to partition the map, combines the l diversity and Laplace mechanisms to filter out k-1 dummy locations, which constitutes k-anonymity with the real locations, and employs an indexing mechanism to protect the user's query privacy. reference [8] proposed a Dummy Location Selection (DLS) algorithm based on entropy metric, which constructs a set of dummy locations with the same query probability to confuse the attacker through the background information of historical query probability of the user's location, and ultimately protects the user's location privacy effectively.

In summary, with the increasing diversity of privacy attacks, attackers may have multiple background knowledge at the same time, and the privacy protection provided by dummy locations is quite fragile if a sound selection strategy is not employed when selecting dummy locations. Existing schemes are by no means designed to ensure that the attacker has one type of background knowledge for selecting dummy locations, and in scenarios where the attacker has composite background knowledge, there is a high probability that these dummy locations will be identified. And although the existing dummy location selection algorithms include the historical query probability and location semantics for location privacy protection, they do not consider the continuous time background of

the same location unit user access to the trend of changing a large difference in the user's privacy protection strength is not high enough. In this paper, we design a dummy location generation method based on spatiotemporal correlation.

2.2 Relative definitions

Definition 1. Location Semantics.

A location that contains features such as latitude, longitude, and semantic location type. Attribute characteristics include:

a) User-defined sensitivity of the semantic location.

b) Multiple locations can have the same semantics, and one location can have multiple semantics.

Definition 2. Euclidean distance.

Indicates the distance from a location $L_i(x_i,\ y_i)$ to other historical location $L_j(x_i,\ y_i)$, where, x , y denotes the latitude and longitude of the location unit. The specific calculation formula is:

$$dis(L_i, L_j) = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{1}$$

Definition 3. Historical query probability.

The area around the user is divided into grids, each grid represents a location cell, and the query probability for each location cell is calculated with the following formula:

$$P_i = \frac{N_i}{\sum_{i=1}^{n} N_i} \tag{2}$$

where: N_i denotes the number of queries for the historical location cells of the grid; $\sum_{i=1}^{n} N$ denotes the sum of queries for all *i*=l historical location cells.

Definition 4. Locational dispersion.

Locational dispersion is a metric used to measure the dispersion of the distribution of locational points between dummy locations and is evaluated in terms of the minimum distance between dummy locations and the area of the polygonal anonymized region formed by the locational points in the pseudo-set, respectively. The minimum distance was calculated using the Euclidean distance as the metric and the polygonal anonymized region area was calculated using the Helen formula. The smaller the minimum distance and the anonymized area are, the easier it is for the attacker to lock the user's real location based on the distribution of the dummy locations; conversely, the more difficult it is for the attacker to determine the user's real location, and thus the more the user's location privacy protection can be strengthened.

Definition 5. Spatial sensitivity.

Spatial sensitivity is used to describe the degree of similarity in the trend of the respective user visits between two location units at different time periods within a day. Mapping the visits of M*M location units within the query geographic location at different time periods within a day's time, the Spatial Sensitivity Matrix (SSM, Spatial Sensitivity Matrix) is constructed, i.e:

$$\begin{bmatrix} \mathbf{A}_1^{\mathbf{T}_1} & \mathbf{A}_1^{\mathbf{T}_2} & \cdots & \mathbf{A}_1^{\mathbf{T}_n} \\ \mathbf{A}_2^{\mathbf{T}_1} & \mathbf{A}_2^{\mathbf{T}_2} & \cdots & \mathbf{A}_2^{\mathbf{T}_n} \\ \vdots & \vdots & \vdots & \vdots \\ \mathbf{A}_{\mathbf{M}^2}^{\mathbf{T}_1} & \mathbf{A}_{\mathbf{M}^2}^{\mathbf{T}_2} & \cdots & \mathbf{A}_{\mathbf{M}^2}^{\mathbf{T}_n} \end{bmatrix} \tag{3}$$

Where Tj denotes the jth time period of a day, Ai denotes the number of user accesses of the ith location unit in the jth time period The spatial sensitivity between the two location units is based on the spatial sensitivity matrix and adopts cosine similarity as a measure, combining with the generated spatial sensitivity matrix, the higher the spatial sensitivity between the two location units, the closer the trend of the number of user accesses of the two location units at different times of a day is. The higher the spatial sensitivity between the two location units, the closer the trend of the number of user visits in different time periods in a day.

3. The Algorithm

This paper proposes a spatiotemporal correlation-based dummy location generation method, utilizing a greedy algorithm and an iterative selection and evaluation strategy, aimed at enhancing location privacy protection. The main ideas can be divided into four steps:

Firstly, a candidate set is generated based on the user's real location and the time when the user initiates the location privacy protection request. Specifically, by analyzing the user's historical query behaviors within a certain past time, 1k points that have the closest query probability to the current real location are identified. These points are considered as the candidate set. They are highly correlated with the user's real location in both time and space, effectively misleading potential attackers. The key to this step is leveraging historical data and probability analysis to ensure that the points in the candidate set have a high access frequency within the specified period, thereby enhancing the depth and breadth of privacy protection.

Secondly, the expected values of spatial sensitivity and semantic similarity between each location point in the candidate set and the user's real location are calculated. Spatial sensitivity refers to the proximity of two locations in the geographic space, which can be measured by geographic distance or transportation convenience. Semantic similarity measures the similarity of two locations in terms of user behaviors patterns, such as whether both locations are commercial areas or residential areas, and whether they have similar environmental characteristics. By combining the expected values of these two dimensions, it ensures that the selected dummy locations are not only geographically close to the real location but also have a high similarity in user behaviours patterns, further enhancing privacy protection. This step employs precise mathematical calculations and modelling to make the selection of dummy locations more scientific and reasonable.

Then, during the selection of dummy locations, it is necessary to maximize the expected value of semantic-spatial sensitivity while ensuring that the points included in the dummy location set can generate the largest anonymous area. The anonymous area size is an important indicator of the strength of privacy protection; the larger the area, the more difficult it is for attackers to infer the user's real location. Therefore, when selecting dummy locations, one must consider not only the similarity between points but also the coverage of the entire anonymous area. By balancing these two aspects, the optimal privacy protection effect can be achieved. The specific selection strategy can be implemented through a greedy algorithm, where each time an optimal point is selected to join the dummy location set, the selection strategy for the remaining points is updated, gradually constructing an optimal dummy location set.

Finally, a dummy location set SDummy, containing the user's real location and with a size of k, is generated. This dummy location set not only meets the requirements of spatial sensitivity and semantic similarity but also ensures the maximization of the anonymous area, thereby effectively enhancing the level of location privacy protection. This method can provide stronger location privacy protection in a dynamically changing spatiotemporal environment. The generation of this dummy location set needs to consider not only the current protection needs but also potential future privacy leakage risks. Through dynamic adjustment and updating, it ensures that users can obtain continuous

and effective privacy protection in different temporal and spatial scenarios.

4. Conclusion

In summary, the method proposed in this paper significantly enhances location privacy protection through meticulous spatiotemporal analysis and optimized selection strategies. This method not only demonstrates its superiority at the theoretical level but also proves its feasibility and effectiveness in practice. By conducting an in-depth analysis of users' historical location data and precisely calculating dummy locations, the proposed strategy can significantly improve location privacy protection without significantly increasing computational overhead. In practical applications, this method is highly flexible and can be further optimized and refined according to specific scenario requirements. For instance, in different application environments, users may have varying privacy protection needs and behaviours patterns. The proposed method can be adjusted based on these specific requirements to provide more personalized privacy protection services. By dynamically adjusting the dummy location generation strategy, this method can adapt to various complex spatiotemporal scenarios, ensuring users receive continuous and effective privacy protection across diverse usage environments. Moreover, this method considers both spatial sensitivity and semantic similarity during the generation of anonymous regions, achieving efficient privacy protection.

In conclusion, the spatiotemporal correlation-based dummy location generation method proposed in this paper holds significant theoretical importance, providing a new research perspective and methodology in the field of location privacy protection. Additionally, it showcases outstanding application potential in practice. By further optimizing and refining the method based on specific application scenarios, it promises to offer users more secure and personalized privacy protection services in the future, becoming an important driving force in the development of location privacy protection technologies. The proposal and application of this method not only enrich the theoretical foundation of location privacy protection but also provide new ideas and tools for practical privacy protection, offering valuable insights and references for the future development of related technologies.

Acknowledgement

This present research work was supported by Harbin Normal University Higher Education Teaching Reform Research Project (No. XJGZ202409).

References

[1] Xiong Jinbo, Bi Renwan, Tian Youliang, et al. Mobile swarm intelligence for sensing security and privacy:models, progress and trends[J]. Journal of Computing. 2021, 44(9):1949-1966.
[2] Li Weihao, Cao Jin, Li Hui, A privacy protection scheme based on location service privacy self-association[J]. Journal of Communication, 2019, 40(5):57-66.
[3] Pan Xiao, Hao Xing, Meng Xiaofeng, Research on continuous query privacy protection in location-based services[J]. Computer Research and Development, 2010, 47(1):121-129.
[4] Xing Ling, Jia Xiaofan, Zhao Pengcheng, et al. A matrix encryption based privacy protection method for UAV swarm location[J]. Journal of Aeronautics, 2022, 43(8):325386-325388.
[5] Jiang Haiyang, Zeng Jianqiu, Han Ke, et al. Research on location privacy protection method for mobile users in 5G environment[J]. Journal of Beijing Institute of Technology, 2021, 41(1):84-92.
[6] Jiao Zexin, Zhang Lin, Liu Xiping. A fine-grained vacation location selection algorithm with differentiated time periods[J]. Journal of Nanjing University of Posts and Telecommunications: Natural Science Edition, 2022, 42(6):106-114.
[7] Jie Wang, Chunru Wang, Jianfeng Ma, et al. False location selection algorithm based on location semantics and query probability[J]. Journal of Communication, 2020, 41(3):53-61.
[8] Liang Huichao, Wang Bin, Cui Ningning, et al. Privacy-preserving approach for point-of-interest queries in road network environment[J]. Journal of Software, 2017, 29(3):703-720.