Research on Location Privacy Protection of Dynamic Anonymous Domain Based on Grid User Density

Ya-lin MIAO, Huan-huan JIA, Yang ZHANG, Xue-min LIU and Tian-tian JI
Xi’an University of Technology, Xi’an, China

Keywords: Privacy protection, Anonymous location, Grid user density.

Abstract. The existing location anonymity algorithms do not consider the distribution of user density in the region. The area of anonymous domain is not the most appropriate and the query workload is redundant. To solve this problem, this paper proposes location privacy protection method for dynamic anonymous domain based on grid user density, which can shrink and expand anonymous domain reasonably on the basis of regional user density. The experimental results show that the proposed algorithm can find smaller anonymous areas under the premise of satisfying user privacy configuration, and effectively compress processing time, improve the success rate of anonymity and LBS service quality.

Introduction

With the popularity of smart mobile devices and high-resolution spatio-temporal data sensors, spatial location technology has been greatly developed. In recent years, based on spatial positioning technology, the application of location based services (LBS) [1] technology using mobile Internet as transmission medium has shown an explosive growth trend. The application of LBS technology is rich and varied, and the integration of "LBS+" and multi-industry development provides more convenient services for people's clothing, food, housing and transportation, which is widely welcomed by users [2,3]. Although the emergence and development of LBS services have many potential benefits, it also opens the door for users' privacy threats.

In order to provide users with more secure, reliable, efficient and convenient LBS services, various methods of location privacy protection have been proposed. In 2003, Gruteser et al. first proposed the core concept of location privacy protection based on k-anonymity [4]. LBS queries were initiated by using a set of location information containing at least k user nodes instead of the user's real location, which reduced the probability that the user's location was identified by the attacker to less than \(1/k\). IC algorithm [5] is proposed at the same time. In geographic space, this algorithm uses quadtree structure to divide spatial region information, and generates anonymous regions including at least \(k\) users to replace the real location of users for LBS service query. Casper Cloak improves the algorithm of generating anonymous domain on the basis of IC algorithm, and proposes Casper algorithm [6]. It uses third-party anonymity architecture to manage spatial index information of quadtree. The algorithm of searching anonymous region is more meticulous, and the rectangular anonymous space obtained is smaller than IC algorithm.

Later, the Grid-divide method, which uses fixed grid index space and dynamically expanding grid to find anonymous domain, was proposed [7,8]. It can obtain smaller anonymous space, but the efficiency of the algorithm needs to be improved. In 2007, Hilber Cloak [9] introduced Hilbert curve to construct anonymous region, mapped two-dimensional space into one-dimensional space and divided the set. Finally, the region where the set satisfies the privacy level was selected as the anonymous domain. NNC and Lique Cloak [10] transform the search of anonymous regions into set search problem, and transform the area of regions into the minimum number of point clusters for calculation. In 2006, Chow CY proposed a P2P location privacy protection scheme [11], which delineated a new distributed P2P location protection architecture. In 2011, Coprivacy [12] proposed an improved scheme based on the SpaceTwist scheme. It used collaboration among multiple user nodes to complete anonymity in an ad hoc network, and selected proxy anchors to initiate LBS service.
requests. Finally, incremental proximity query method was used to refine the query results according to the distance between users and proxy points. In 2013, Chen Yufeng et al. [13] proposed a location privacy protection method based on game theory based on the cooperative nodes of untrusted users under P2P structure, which further improved the efficiency of algorithm query.

Inspired by Casper's algorithm, this algorithm uses quadtree grid index model. Considering the distribution of users in the region, it is not necessarily able to find the most suitable anonymous domain in the case of uneven user density distribution. To solve this problem, this paper proposes location privacy protection method for dynamic anonymous domain based on grid user density (GDOA), which dynamically searches for the most suitable anonymous domain, reduces the anonymous area as much as possible, reduces the redundant query workload, and thus improves the quality of LBS service.

Related Work

Based on the general third-party anonymity architecture, this paper designs the system model structure. The main components of the system include user location terminal, trusted central anonymity device and LBS server database. The LBS requests of the user location terminal need to be sent to the LBS server anonymously through the central anonymity device, and then returned to the mobile user terminal after the result refinement by the central anonymity device.

Spatial Grid Index Model

Quadtree (Q-Tree) is a commonly used technology of spatial index in GIS. It is also widely used in image processing, game collision detection and so on [14]. As shown in Fig.1(c), Q-Tree is distributed in a tree structure, with root node at the top and leaf node at the bottom. In addition to leaf node, each node points to four child nodes respectively. In addition to root node, each child node only points to one father node, so as to establish a data quadtree model according to the index rule.

In spatial indexing, a geographic area covered by a two-dimensional rectangle is usually regarded as the root node. As shown in Fig.1(b), the root node area is divided into four quadrants or four rectangular grid areas after a partition. Each generated grid area is added with an index identifier and is recorded in the grid identification hash table as shown in Fig.1(a), and the relevant user information in the grid is stored in the user hash table of corresponding node in the database. In the location privacy protection method, a grid spatial index model based on Q-Tree is used to index and store mobile user information in the spatial area.

In this paper, grid partitioning algorithm based on the user density, first uses Q-Tree spatial index model to partition the spatial area, manages the spatial grid area by index, calculates the user distribution density in each grid area, and designs the database storage structure of anonymity device to achieve efficient storage and management of geographic information and user information.

Location Privacy Protection Algorithm

When users first make LBS requests, they need to personalize their privacy configuration information in an anonymous device, while the anonymous device updates and maintains the database accordingly. When users make snapshot queries, the anonymity device uses quadtree algorithm model to mesh the jurisdiction area, updates users' location in the grid and grid-related data information in real time,
anonymizes users' location according to the privacy profile registered in real time, and generates anonymous areas conforming to the privacy configuration. The specific steps of the algorithm are as follows:

Step 1: Traversing the data table, querying the number $rid$ of the smallest grid in which user $i$ is located, obtaining the information $G_i$ of the grid, indexing the user privacy configuration table to get the user privacy configuration $P_i$, judging whether the number of mobile users in the grid meets the anonymity degree, and whether the grid area meets the privacy configuration, satisfying the direct output grid as the anonymous domain $c_i$, if not, then execute the next step.

Step 2: Quadtree index obtains that the parent grid of the smallest grid is identified as $pid$, and the information of the parent grid is obtained by $pid$ index, which is recursive to the $h$ layer, so that the $G_i$ of $rid(h)$ can be found to satisfy the user's privacy configuration $P_i$, so that $rid = rid(h-1)$, and then the next step is executed.

Step 3: A circular contraction anonymous domain generation algorithm is implemented for the grid marked $rid$ at this time. If the generated anonymous domain satisfies the privacy configuration $P_i$, it returns to the circular anonymous domain $c_i$, and if it is not satisfied, the next step is executed.

Step 4: Implement the grid user density expansion anonymity algorithm for $rid$ grid. The user density of four adjacent grids in vertical and horizontal directions is calculated and expanded to the grid with high user density. At this time, the area of anonymous domain is twice that of $rid$ grid. It is judged whether the anonymous domain satisfies the privacy configuration $P_i$ and returns to the anonymous domain $c_i$, otherwise the next step will be executed.

Step 5: Return directly to the grid $rid(h)$ of step 2, which meets the privacy requirement, and return it as the anonymous domain $c_i$. If the above steps cannot construct the anonymous domain that meets the requirement, or the anonymous time exceeds the maximum user tolerance time $\Delta T$, the anonymous domain construction will fail.

When the generation of anonymous domain is successful, the information of anonymous domain is recorded, and the anonymous domain is sent to the location service provider LSP along with the user's LBS query content. When the generation of anonymous domain fails, the failure mark is returned to the user.

Fig.2 depicts an example of algorithm execution when anonymity $k$ is different. Fig.2(a) is a meshed geographic area model. Solid triangle represents the user $i$ that initiates LBS queries, and solid dots represent the distribution of other users. When $k = 5$, the LBS query is sent to the user, and only one user in the grid does not conform to $P_i$. Executing the algorithm, as shown in Fig.2(b), the algorithm recursively queries twice. The number of users in the grid is 3 and 20 respectively. The grid with 3 users does not satisfy the degree of anonymity. As shown in Fig.2(c), the number of users in the area covered by the outer circle of the grid with the center of the grid is calculated by using the circular contraction algorithm. Two solid pentagonal users are visible between the grid and the outer circle, so that the number of users in the outer circle area reaches 5, satisfying the degree of anonymity, and then they can return to the area covered by the outer circle as the ultimate anonymous domain $c_i$. When $k = 10$, the anonymous algorithm of grid user density expansion is implemented to calculate the user density of four vertical and horizontal adjacent grids, respectively, and expand to the direction of high user density. As shown in Fig.2(d), the area is expanded to twice the area of the grid, and the number of users of the shaded rectangular area is 11, which satisfies the degree of anonymity, so the rectangular area is returned as the anonymity domain $c_i$. When $k = 15$, as shown in Fig.2(e), the number of users in the shadow grid is 20, which satisfies the degree of anonymity $k$, so the shadow grid is returned as the anonymity domain $c_i$. 

222
It is not difficult to see from the example that the algorithm in this paper refines the generation of anonymous domain into several levels according to different levels of privacy to ensure that the more suitable anonymous domain can be returned as far as possible, and the algorithm complexity is less than Casper algorithm, and the anonymity time is less.

Experiments

The development platform of this algorithm is Intel (R) Core (TM) i5-2430M CPU @ 2.40GHz, 6GB, x64 Win10 operating system. Using Java programming language, the algorithm is implemented and simulated on the software system platform built by Eclipse and MySQL. A data generator TNGMO platform [15] based on mobile objects of road network developed by Thomas Brinkhoff team in Germany was selected to generate experimental simulation data sets. The mobile objects generated in a rectangular area of about 16000*16000 in the traffic network of OlenburgGen, Germany, are used to simulate the mobile users in the real life logistics and transportation environment. Quadtree algorithm model is used to divide and index the regional space, and the spatio-temporal data sets of mobile objects with different user densities are selected from the generated data sets for simulation and performance evaluation.

In this paper, GDOA algorithm, IC algorithm and Casper algorithm are respectively constructed and programmed. The performance experiments of the three algorithms are carried out under the same hardware and software configuration. And select the anonymous success rate and effective time index commonly used in location anonymous domain algorithm to compare and evaluate the algorithm.

The success rate of anonymity is a measure to show the effectiveness of the algorithm intuitively. In this paper, k anonymity is divided into 10 equal values in [5,50]. Mobile object datasets with user density of 2500 are selected. Maximum user tolerance time is specified in each snapshot query. Success rates of constructing anonymous domains that meet users' privacy requirements are counted. As shown in Fig.3, the anonymity success rate of GDOA algorithm is significantly higher than that of the other two algorithms, and the success rate is stable at about 90%. This shows the stability of constructing anonymous domains that meet users' privacy needs in user tolerance time. The size of regional user density will affect the efficiency of anonymous domain generation. This paper simulates
the algorithm based on different user density 200, 1000, 1500, 2000 and 2500. The experimental results are shown in Fig.4. Under different user densities, the anonymity success rate of GDOA algorithm is higher than that of other two algorithms, and the success rate of GDOA algorithm increases steadily with the increase of user density.

Figure 3. Effect of anonymity on anonymous success rate.    Figure 4. Effect of user density on anonymous success rate.

Effective time refers to the time consumed by the user in the process of successfully obtaining corresponding location services from LSP. As shown in Fig.5, the average anonymity time of the three algorithms increases significantly with the increase of $k$ anonymity. The average anonymity time of the algorithm in this paper is the lowest and the efficiency of the algorithm is the highest compared with the other two algorithms. As shown in Fig.6, with the increase of user density, the anonymity time consumed by all algorithms increases with the increase of the number of requests sent by users. Because of the highest complexity of the algorithm, Casper algorithm takes the longest time. When the user density is less than 1500, the anonymity time of IC algorithm is more advantageous, but with the increase of the number of users, the algorithm in this paper is more efficient and stable than the other two algorithms in terms of effective time.

Figure 5. Effect of $k$ anonymity on effective processing time.                Figure 6. Effect of user density on effective time.

Based on the above experimental results, the proposed algorithm can find the smallest anonymous area as possible under the premise of satisfying user privacy configuration, and effectively compress processing time to improve the success rate of anonymity and the quality of LBS service.

Conclusion

Based on the rule of user distribution in an area as the premise of anonymous domain generation, this paper proposes location privacy protection method for dynamic anonymous domain based on grid user density for the purpose of generating the optimal anonymous domain. Through the comparative analysis of the experimental results, it is shown that the proposed algorithm shrinks and expands the anonymous domain reasonably on the basis of regional user density, and effectively reduces the area of anonymous domain and improves the quality of service while satisfying the configurations of user privacy parameters.
Acknowledgement
This work was supported by the 2017 Xi'an Science and Technology Project, number: 2017050NC/NY011(1). We thank senior engineer Yilong Xiao for excellent technical support and valueable discussion.

References
[1] Tooska Dargahi, Moreno Ambrosin, Mauro Conti, et al. ABAKA: A novel attribute-based k-anonymous collaborative solution for LBSs [J]. Computer Communications, 2016, 85:1-13.
[2] Wu Sheng, Wang Qinmin, Xiao Guirong, et al. Application and framework of LBS in ITS [J]. Journal of Surveying and Mapping Science and Technology, 2006, 23(1): 11-14.
[3] Zhou Aoying, Yang Bin, Jin Chengqing, et al. Location-based services: architecture and progress [J]. Journal of Computer Science, 2011, 34 (7): 1155-1171.
[4] Fuyu Liu, Kien A. Hua, Ying Cai. Query l-diversity in Location Based Services[C]. International Conference on Mobile Data Management: Systems Services and Middleware, 2009:436-442.
[5] Marco Gruteser, Dirk Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking[C]. International Conference on Mobile Systems, Applications, and Services DBLP, 2003: 31–42.
[6] Mokbel M.F, Chow Chi-Yin, Aref W.G. The New Casper: Query Processing for Location Services without Compromising Privacy[C]. International Conference on Very Large Data Bases, 2006: 763-774.
[7] Zou Yonggui, Zhang Yuhan. Location anonymity algorithm based on grid partition space [J]. Computer application research, 2012, 28 (9):3059-3061.
[8] Wang Weifang, Chen Mingrui, Zhao Yaochi. An improved spatial location anonymity algorithm based on grid partition [J]. Computer simulation, 2016, 33 (9): 208-210.
[9] Panos Kalnis, Gabriel Ghinita, Kyriako Mouratidis, et al. Preventing Location-Based Identity Inference in Anonymous Spatial Queries [J]. IEEE Transactions on Knowledge & Data Engineering, 2007, 19(12): 1719-1733.
[10] Bugra Gedik, Ling Liu. Protecting Location Privacy with Personalized k-Anonymity: Architecture and Algorithms [J]. IEEE Transactions on Mobile Computing, 2008, 7(1): 1-18.
[11] Chi-Yin Chow, Mohamed F. Mokbel, Xuan Liu. A Peer-to-Peer Spatial Cloaking Algorithm for Anonymous Location-based Services[C]. Proceedings of the 14th Annual ACM International Symposium on Advances in Geographic Information Systems, ACM, 2006: 171-178.
[12] Huang Yi, Huo E, Meng Xiao Feng. Co Privacy: A User Collaboration Location Privacy Protection Method for Anonymous Area [J]. Computer Journal, 2011, 34 (10): 1976-1985.
[13] Chen Yufeng, Liu Xuejun, Li Bin. User Phase Based on Game Theory Collaborative Location Privacy Protection Method [J]. Computer Science, 2013, 40 (10): 92-97.
[14] Li Jianxun, Shen Bing, Jiang Rengui, et al. Quadtree spatial indexing algorithm for image pyramid [J]. Computer Engineering, 2011, 37 (10): 11-13.
[15] Thomas Brinkhoff. A Framework for Generating Network-Based Moving Objects[J]. Geoinformatica, 2002, 6(2): 153-180.