Protecting Location Privacy Through Crowd Collaboration

Zhonghui Wang, Guangwei Bai(✉), and Hang Shen

College of Computer Science and Technology,
Nanjing Tech University, Nanjing 211816, China
bai@njtech.edu.cn

Abstract. Location-based services (LBSs) enable users to sense their surroundings at the risk of exposing coordinates to attackers. Worse yet, a strong adversary with arbitrary knowledges probably derive more privacy especially in continuous query scenarios. To address the problems, a multi-player privacy game mechanism is proposed to satisfy users' location privacy against adaptive attacks while maximizing utility, building upon which a heuristic algorithm is applied to iteratively converge to the optimal equilibrium point. The gain stems from the collaboration of mobile devices: users share information and forward queries for each other. We evaluate our mechanism against the Bayesian localization attack and maximum possible moving speed attack. The simulations with real map data and mobility traces indicate that our mechanism is effective to preserve privacy at an acceptable price of utility and time complexity.

Keywords: Location-based service · Multi-player privacy game · Joint differential-distortion privacy · Inference privacy · Adaptive attack

1 Introduction

Users are enabled to query the LBS servers for the purpose of searching points of interest (POIs, like restaurants, stores, etc.), real-time traffic information or navigation related to the current position, which is observable to attackers [1,2]. Sensitive coordinates may be exposed during the queries. Even worse, the strong adversary [3,4] with arbitrary knowledges probably traces and models the queries to predict users' following behaviors and derive more privacy.

Data confusion is an excellent mechanism for hiding sensitive data by misleading, extra or ambiguous information, resulting in extra extracting overhead. A number of obfuscation mechanisms have been proposed [5,6]. One of the most important is Stackelberg Game proposed in [7], where the focus is on two rivals solved by linear programming. Another important contribution is joint differential-distortion privacy metric. The privacy achieved through joint metric against optimal attacks is the maximum privacy that can be achieved by either of these metrics separately. The utility cost is also not larger than what

Q. Chen et al. (Eds.): ChinaCom 2016, Part I, LNICST 209, pp. 179–188, 2018.
DOI: 10.1007/978-3-319-66625-9_18

either of them imposes. However, it fails to preserve privacy in continuous query scenarios. When observing queries continuously issued, the adversary may run overlapping rectangle attack [9], continuous query attack [10] or maximum possible moving speed attack [11] by linking historical cloaking regions with users' mobility patterns to infer more privacy than obtained from an isolated query.

Even worse, cooperations between the adversary and LBS providers greatly weaken users' privacy. Fortunately, with the rapid advance in mobile devices and their embedded sensors, users in local area can help each other to enhance privacy protection without a trusted central server [8]. We enable users to randomly select neighbors for forwarding queries through a transition probability matrix P. Thus users are supposed to negotiate with each other over P so that more privacy can be preserved.

To address the above concerns, we propose a multi-palyer game mechanism to minimizing users' utility loss with respect to privacy measured by both inference and joint differential-distortion privacy metrics, where the adversary runs adaptive attacks to minimize users' privacy by inverting users' strategies. On the basis, a heuristic algorithm is proposed to iteratively converge to the optimal equilibrium point on P. The simulations with real map data and mobility traces indicate that our mechanism is effective to preserve privacy at an acceptable price of utility and time complexity. Additionly, compared with the existing joint differential-distortion privacy metric, employing both inference and joint privacy metrics significantly improves the privacy level.

The remainder of this paper is organized as follows. Section 2 defines some important concepts and states problems. In Sect. 3, an optimal obfuscation mechanism is designed for multi-palyer game scenarios, followed by thorough analysis and evaluation in Sect. 4. Finally, Sect. 5 concludes the paper.

2 Problem Statement

We assume a user wants to protect sensitive information when communicating with untrusted LBS providers, and refer to his sensitive information as *secret*, which can be protected by collaboration. More specifically, the user may asks others (assuming there's no spiteful users) to issue LBS queries for him according to P. Figure 1 illustrates the information flow. The joint probability distribution

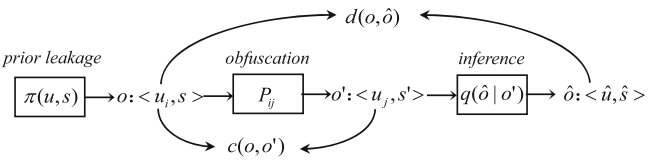


Fig. 1. The information sharing framework. A user-secret pair $<u, s>$ denoted by o is obfuscated into observable o' by the mechanism according to P. The adaptive adversary runs inference attack q on o' and draws a probability distribution over estimates $\hat{o}$. Distance function d stands for privacy level, and c denotes the utility cost.

π is estimated by observing the users' exposed information in the past. Thus we need to update π whenever any user shares his information. In the following sections, we define several important concepts and state the problems based on the information sharing framework.

2.1 Weighed Distance

In fact, a user could bear much smaller distance error if he is at a bus stop or other POIs with less sensitivity. In order to reduce unnecessary computational cost, we provide adjustable protection level.

Thus the notion of weighed distance is introduced to provide flexibility. POIs are classified into several levels from extremely sensitive to not sensitive. The coefficient of ith level is w_i, which is defined by users before entering in or modified during the game. The weighed distance is calculated as

$$d^{w_i} = w_i \cdot d, w_i \in [0, 1] \tag{1}$$

The smaller w_i is, the more sensitive related POIs are.

2.2 Joint Differential-Distortion Privacy Metric

After an observable o' was released, the adversary will speculate about the original content of the query and get an estimate $\hat{o}$. Therefore we use the distance between o and $\hat{o}$ to define distortion privacy. A user would be less worried about revealing $o \sim p\,(o'|o)$, if the portrait of his secret o in the eyes of the adversary is an estimate $\hat{o}$ with a large distance $d^w(o, \hat{o})$.

Given inference algorithm q and specific secret o, the user's privacy obtained through a protection mechanism p is computed by

$$\sum_{o'} p(o'|o) \sum_{\hat{o}} q(\hat{o}|o') d^w(o, \hat{o}). \tag{2}$$

The expected distortion privacy of the users is

$$\sum_{o} \pi(o) \sum_{o'} p(o'|o) \sum_{\hat{o}} q(\hat{o}|o') d^w(o, \hat{o}). \tag{3}$$

A generic definition of differential privacy is adopted in this paper, assuming arbitrary distance function d^ϵ on the users. A protection mechanism is differentially private if for all users $u, u' \in U$ with distinguishability $d^\epsilon(u, u')$, and for all observables $o' \in O$, we have

$$p\,(o'|\,\langle u, s\rangle) \leqslant \epsilon d^\epsilon\,(u, u') \cdot p\,(o'|\,\langle u', s'\rangle)\,. \tag{4}$$

In fact, the differential privacy metric guarantees that, given the observation, there is not enough convincing evidence to prefer to one user than others.

2.3 Inference Privacy Metric

Users usually continuously issue queries. The adversary with background knowledges can infer more information aftering comparing adjacent queries. Given that the protection mechanism p releases o'_t at time t and last output is o'_{t-1}, ε-privacy can be saved if the following inequality holds.

$$q\left(\hat{o_t}|o'_t, o'_{t-1}\right) \leqslant e^{\varepsilon} q\left(\hat{o_t}|o'_t\right), \tag{5}$$

where more privacy is preserved when ε getting smaller.

2.4 Utility Cost

Through obfuscation mechanism, users may gain more privacy while incurring more utility loss. On one hand, it leads to additional communication cost denoted by c_{d} when user i asks j to forward the query instead of issuing it personally. On the other hand, we explain the heterogeneity between two secrects by an r-range query (considering users will most likely issue r-range queries or kNN queries and both of them are related to circle regions). Figure 2 describes how j handles the query from i. The mechanism increases the cost of data transmission and the workload of the result refinement process. All these expenditure are proportional to the size of superset, depending on the quantity of POIs within this range. Because there is locally even distribution in POIs, we calculate the extra refinement cost c_{r} with density and area instead of the quantity of POIs.

$$c_{\mathrm{r}} = \frac{\rho \cdot \pi \cdot r'^2}{\rho \cdot \pi \cdot r^2} \geqslant \frac{(r + d_{ij})^2}{r^2}, \tag{6}$$

where ρ is density of POIs and d_{ij} is the linear distance between u_i and u_j.

To simplify the problem, we employee function c to denote the overall utility cost calculated by c_{d} and c_{r}.

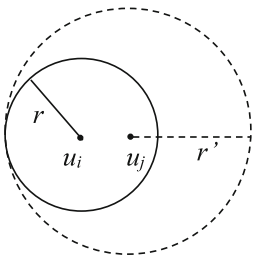


Fig. 2. Processing procedure of a range query. Firstly, u_i issues a range query with radius r and selects u_j to forward the query. Aftering receiving the original request o, u_j computes a new radius r' and repacks it into o'. Then LBS providers will return a result superset after processing o'. Finally, u_j needs to extract the exact results what u_i requests and send it back to finish the entire procedure.

2.5 Objective

The objective is to find the optimal balance between privacy and utility, and to construct the protecting mechanism that achieves such an equilibrium point. In other words, it is to find a vector of probability distribution function p^* to minimize the overall utility cost, on average,

$$p^* = \min \sum_i \pi(u_i) \sum_{j,s} \pi(u_i, s) P_{ij} c_{ij} \tag{7}$$

under users' privacy constraints.

Let d_m be the minimum desired distortion privacy level and ϵ_m be the differential privacy budget associated with the minimum desired privacy of the users. The users' joint privacy is guaranteed if p satisfies

$$\sum_{o'} p(o'|o) \sum_{\hat{o}} q^*(\hat{o}|o') d^w(o, \hat{o}) \geqslant d_m, \forall o \in O, \tag{8}$$

$$p\left(o' | \langle u, s \rangle\right) \leqslant \epsilon_m d^{\epsilon}\left(u, u'\right) p\left(o' | \langle u', s' \rangle\right), \forall u, u', o'. \tag{9}$$

Let ε_m be the desired inference privacy of the users. The users' inference privacy is guaranteed if

$$q^*\left(\hat{o_t} | o'_t, o'_{t-1}\right) \leqslant e^{\varepsilon_m} q^*\left(\hat{o_t} | o'_t\right), \forall t. \tag{10}$$

With the objective, the following multi-player game minimizes the overall cost when satisfying the above constraints.

3 Privacy Game

Definition 1 *(Privacy Game). A strategic game consists of*

1. *A finite set M: the set of players,*
2. *For each player $i \in M$, a nonempty set A_i: the set of actions available to player i,*
3. *For each player $i \in M$, a preference relation $\succeq_i$ on A.*

$\succeq_i$ is defined by a utility cost function c_i. For any $a \in M, b \in M$, $c_i(a) \leqslant c_i(b)$ if $a \succeq_i b$.

Each player wants to maximize the objective according to his preference relation. A user's action space is all users he can request for forwarding queries. The adversary's action space is all possible requesters when observing outcomes of obfuscation mechanism. Assuming that the obfuscation mechanism is not oblivious and is available to all players, the adversary takes the upper hand in the conflict for making decisions after users. Therefore, an obfuscation mechanism against a fixed attack is always suboptimal. The best obfuscation mechanism should be designed against any adaptive attack which is tailored to each specific obfuscation mechanism. After the adversary designs the best inference attack,

users' goal is the obfuscation against the adversary. Accordingly, we do not model any particular adversary but the one who minimizes users' privacy according to observation.

Given secret o, we denote a mixed strategy for user u_i by

$$\begin{aligned} p_i =& p\left(\cdot|o\right) = \left\{p\left(o'_1|o\right), p\left(o'_2|o\right), ..., p\left(o'_j|o\right), ...\right\}, \\ & \forall o'_j \in O, p\left(o'_j|o\right) \geqslant 0, \text{and} \sum_j p\left(o'_j|o\right) = 1. \end{aligned} \tag{11}$$

Similarly, let q be the set of the adversary's mixed strategy of finding out the original requester when observing o',

$$\begin{aligned} q =& q\left(\cdot|o'\right) = \left\{q\left(\hat{o}_1|o'\right), q\left(\hat{o}_2|o'\right), ..., q\left(\hat{o}_j|o'\right), ...\right\}, \\ & \forall \hat{o}_j \in O, q\left(\hat{o}_j|o'\right) \geqslant 0, \text{and} \sum_j q\left(\hat{o}_j|o'\right) = 1. \end{aligned} \tag{12}$$

p, q and π are available to all players. With these information, users want to figure out the mutually optimal $\langle p^*, q^* \rangle$, which is the solution of the game. In the following sections, we design the optimal attack q^* and the best obfuscation mechanism p_i^* for each user u_i against q^* under his privacy constraints.

3.1 Optimal Strategies

The adversary's objective is to minimize the users' privacy, i.e., to minimize error between the estimation $\hat{o}$ and original secret o. The optimal attack is

$$q^* = \min_q \sum_{\hat{o}} p^*\left(o'|o\right) q\left(\hat{o}|o'\right) d\left(o, \hat{o}\right), \tag{13}$$

where q is not only a Bayesian probability inverse, but also considering mobile pattern attack(MPA) like maximum possible moving speed attack to infer more privacy in continuous query scenarios.

Against the adversary, users cooperate with each other to minimize overall cost under the premise of satisfying every user's privacy. Thus we can formulate the protection as

$$p^* = \min_p \sum_{o,o'} p\left(o'|o\right) q^*\left(\hat{o}|o'\right) c\left(o, \hat{o}\right) \tag{14a}$$

$$\mathbf{s.t.} \sum_{o'} p(o'|o) \sum_{\hat{o}} q^*(\hat{o}|o') d^w(o, \hat{o}) \geqslant d_m, \forall o \in O, \tag{14b}$$

$$p\left(o'|o_i\right) \leqslant \epsilon_m d^{\epsilon}\left(o_i, o_j\right) \cdot p\left(o'|o_j\right), \forall i, j, o', \tag{14c}$$

$$q^*\left(\hat{o}_t|o'_t, o'_{t-1}\right) \leqslant e^{\varepsilon_m} q^*\left(\hat{o}_t|o'_t\right), \forall t. \tag{14d}$$

Equation (14a) is to minimize overall cost of all queries; constraints (14b), (14c) and (14d) represent the desired distortion, differential and inference privacy level of all users.

3.2 Optimal Equilibrium Point

The solution is to find the mutually optimal $\langle p^*, q^* \rangle$ among all pairs. Assuming that there are n users involved in this game, the time complexity of enumerating all $\langle p, q \rangle$ is $O(n^3)$, which is infeasible when n is large enough.

To reduce the complexity, a heuristic algorithm is proposed to iteratively converge to the optimal equilibrium point. Considering hiring a remote user for forwarding queries will take much higher utility cost, users prefer to give closer neighbors a higher forwarding probability. Thus an appropriate initial probability density function is

$$f(x,y) = \frac{1}{\sqrt{2\pi\sigma^2\sqrt{1-\varrho^2}}} e^{-\frac{1}{2\sigma^2(1-\varrho^2)}\left[(x-y)^2+(1-2\varrho)(x-a)(y-b)\right]}, \tag{15}$$

where a, b is the coordinate (usually marked with latitude and longitude in maps) of requester $u(a,b)$, σ and ϱ are parameters for adjusting probability distribution. The neighbors of $u(a,b)$ are ordered by distance as a sequence: $u(x_0,y_0), u(x_1,y_1), ...u(x_i,y_i), u(x_{i+1},y_{i+1})...$ (where $u(x_0,y_0)$ is $u(a,b)$). The probability of chosing u_i, i.e., $u(x_i,y_i)$, is calculated as

$$\iint_D f(x,y)\,dxdy,$$
$$D : (x_i-a)^2 + (y_i-b)^2 \leqslant (x-a)^2 + (y-b)^2 < (x_{i+1}-a)^2 + (y_{i+1}-b)^2. \tag{16}$$

After n users get their own probability transition matrix, the multi-player game has taken the first step to get the initial P before releasing queries like $p(u_j|u_i)$. The adversary runs $q(\hat{u_i}|u_j)$ to find out the original requesters. Afterwards, obfuscation mechanism will modify P by

$$P'(u_j|\hat{u_i}) = \left(1 - \frac{1}{n}\right) q(\hat{u_i}|u_j)\, P(u_j|\hat{u_i}), \tag{17}$$

if u_j' secret is exposed to the adversary. Players repeat the above steps to converge to the equilibrium point, with convergence rate determined by n and q.

4 Performance Evaluation

In this section, the effectiveness of our proposed mechanism are experimentally evaluated under several system settings, with data of Beijing that contains various categories of POIs [12]. To the best of our knowledge, due to privacy and commercial interest reasons, no real suitable data sets have been publicly released. Therefore, in most of our experiments, we randomly generate a group of users as players. In addition, we adapt the real devices data to validate the mechanism. There is no inference attack being specified, but the maximum possible moving speed attack is employed in the experiments to illustrate the crucial problems. The evaluation metrics include privacy level, utility cost and time cost. All algorithms are implemented with Matlab and run on a desktop PC with Intel Core i3 2.53 GHz processor and 8 GB memory.

4.1 Case 1: Impacts of POI Density

This section examines the impacts of POI density, while the other parameters keeping constant. Let a group of users experience LBSs in three different districts, where Haidian is downtown with the highest POIs density, and Yanqing is suburban with the lowest density. We compare the utility cost and privacy of users located in the three districts. In Fig. 3, it is interesting to observe that users can achieve more privacy at lower utility cost when being in Haidian. However, staying in Yanqing probably results in much more expensive cost when satisfying the same privacy level, which means users in districts with higher POIs density can preserve much more privacy with the same cost limitation. That is to say, our proposed mechanism performs better in districts with higher POIs density, though it protects at least 79% privacy in the suburbs.

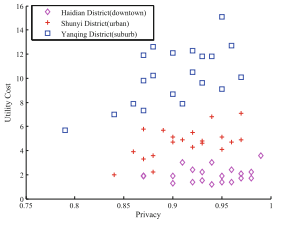


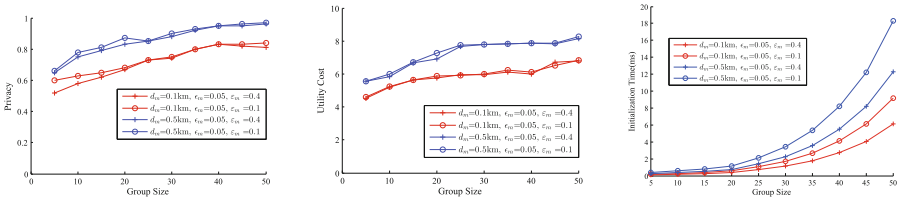
Fig. 3. Impacts of POI density.

4.2 Case 2: Impacts of Privacy Level

In this section, we investigate the impacts of privacy level from two aspects: increasing the average minimum desired distortion privacy level d_m and inference privacy budget ε_m. The differential privacy metric ϵ_m is set as 0.05 for being a static metric with little impacts on indicators. In addition, we analyse the impacts of the group size on indicators, including privacy, utility cost and time cost. Figure 4 depicts impacts of different privacy level restrained by d_m, ϵ_m and ε_m, with POIs data of Shunyi district. Increasing d_m by 0.4 km has much deeper influence than decreasing ε_m by 0.3. We are glad to see the average privacy of 50 users reaches up to 96% when $d_m = 0.5\,\text{km}, \epsilon_m = 0.05$ and $\varepsilon_m = 0.1$. With group size growing from 5 to 50, the average privacy increases by 30%, while utility cost increases at a low speed. Nevertheless, the exponential growth in initialization time raises concerns about choice on group size, though 18 ms might seem to be acceptable. From the analysis, we realize that gathering 20 devices into a group will incur relatively low utility cost and initialization time consuming under privacy constraints.

4.3 Case 3: Tracing Real Devices

As mentioned earlier, no real suitable data sets are publicly available. Thus we develop a software tool to collect trajectory data from 50 real mobile devices and



(a) Average Privacy (b) Utility Cost (c) Time Consuming

Fig. 4. Impacts of privacy level.

take the data as input to evaluate our mechanism. Figure 5 indicates that our obfuscation mechanism can provide a better protection when employing both inference and joint privacy metrics by increasing about 10% privacy. We believe that such a mechanism would be acceptable in terms of privacy level of mobile users.

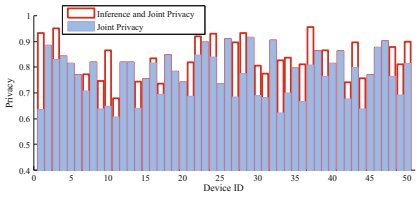


Fig. 5. Tracing data of 50 real devices.

5 Conclusions

This paper contributes to the extensive field of research that concerns obfuscation mechanisms, e.g., in the context of privacy metric, attack algorithm and anonymity in distributed mobile systems without a trusted central server. The proposed obfuscation mechanism is able to preserving users' location privacy against adaptive attacks when maximizing utility. Another important contributions is the optimization with respect to both joint differential-distortion and inference privacy metrics, as well as weighed distance. The simulations with real map and mobility traces corroborate that it is effective to preserve privacy at an acceptable price of utility and time cost. Additionally, it proves that users in districts with higher POIs density can preserve much more privacy with the same cost limitation.

Acknowledgements. The authors gratefully acknowledge the support and financial assistance provided by the National Natural Science Foundation of China under Grant Nos. 61502230 and 61073197, the Natural Science Foundation of Jiangsu Province under Grant No. BK20150960, the Innovation Project for Postgraduates of Jiangsu

Province under Grant No. KYLX16_0600, the Natural Science Foundation of the Jiangsu Higher Education Institutions of China under Grant No. 15KJB520015.

References

1. Hashem, T., Kulik, L., Zhang, R.: Countering overlapping rectangle privacy attack for moving kNN queries. Inf. Syst. **38**(3), 430–453 (2013)
2. Rathi, N., Ghosh, S., Iyengar, A., et al.: Data privacy in non-volatile cache: Challenges, attack models and solutions. In: 2016 Proceedings of 21st IEEE Asia and South Pacific Design Automation Conference, Macao, pp. 348–353 (2016)
3. Wen, F., Li, X.: An improved dynamic ID-based remote user authentication with key agreement scheme. Comput. Electr. Eng. **38**(2), 381–387 (2012)
4. Guo, M., Pissinou, N., Iyengar, S.S.: Pseudonym-based anonymity zone generation for mobile service with strong adversary model. In: 2015 Proceedings of 12th IEEE Annual Consumer Communications and Networking Conference (CCNC), Las Vegas, USA, pp. 335–340 (2015)
5. Garg, S., Gentry, C., Halevi, S., et al.: Candidate indistinguishability obfuscation and functional encryption for all circuits. In: Proceedings of 54th IEEE Annual Symposium on Foundations of Computer Science, Berkeley, USA, pp. 40–49 (2013)
6. Ardagna, C.A., Cremonini, M., di Vimercati, S.D.C., et al.: An obfuscation-based approach for protecting location privacy. IEEE Trans. Dependable Secure Comput. **8**(1), 13–27 (2011)
7. Shokri, R.: Privacy games: optimal user-centric data obfuscation. In: Proceedings of Privacy Enhancing Technologies, Philadelphia, USA, vol. 2015, no. 2, pp. 299–315 (2015)
8. Ghinita, G., Kalnis, P., Skiadopoulos, S.: PRIVE: anonymous location-based queries in distributed mobile systems. In: Proceedings of 16th ACM International Conference on World Wide Web, Banff, Canada, pp. 371–380 (2007)
9. Ma, C., Zhou, C., Yang, S.: A voronoi-based location privacy-preserving method for continuous query in LBS. Int. J. Distrib. Sensor Net. **2015**, 1 (2015)
10. Gustav, Y.H., Wu, X., Ren, Y., Wang, Y., Zhang, F.: Achieving absolute privacy preservation in continuous query road network services. In: Luo, X., Yu, J.X., Li, Z. (eds.) ADMA 2014. LNCS, vol. 8933, pp. 279–292. Springer, Cham (2014). doi:10.1007/978-3-319-14717-8_22
11. Xu, J., Tang, X., Hu, H., Du, J.: Privacy-conscious locationbased queries in mobile environments. IEEE Trans. Parallel Distrib. Syst. **21**(3), 313–326 (2010)
12. POIs Data of Beijing (2016). http://www.datatang.com/data/44484