Location Differential Privacy Protection in Task Allocation for Mobile Crowdsensing Over Road Networks

Mohan Fang(✉), Juan Yu, Jianmin Han, Xin Yao, Hao Peng, Jianfeng Lu, and Ngounou Bernard

Zhejiang Normal University, Jinhua 321004, China
{fangmohan, yujuan, yaoxin, hpeng}@zjnu.edu.cn,
{hanjm, lujianfeng}@zjnu.cn

Abstract. Mobile Crowdsensing (MCS) platforms often require workers to provide their locations for task allocation, which may cause privacy leakage. To protect workers' location privacy, various methods based on location obfuscation have been proposed. MCS over road networks is a practical scenario. However, existing work on location protection and task allocation few considers road networks and the negative effects of location obfuscation. To solve these problems, we propose a Privacy Protection Task Allocation framework (PPTA) over road networks. Firstly, we introduce Geo-Graph-Indistinguishability (GeoGI) to protect workers' location privacy. And then we model a weighted directed graph according to the road network topology and formulate a linear programming to generate an optimal privacy mechanism, which aims to minimize the utility loss caused by location obfuscation under the constraint of GeoGI. We also improve the time-efficiency of the privacy mechanism generation by using a δ-spanner graph. Finally, we design an optimal task allocation scheme based on obfuscated locations via integer programming, which aims to minimize workers' travel distance to task locations. Experimental results on Roma taxi trajectory dataset show that PPTA can reduce average travel distance of workers by up to 23.4% and increase privacy level by up to 21.5% compared to the existing differential privacy methods.

Keywords: Mobile Crowdsensing · Location privacy · Task allocation · Road network · Linear programming

1 Introduction

Mobile Crowdsensing (MCS) [1] is an emerging paradigm that combines of crowdsourcing and mobile devices, it engages workers to collect urban-scale sensing data with sensor-equipped smartphones. MCS has the advantages of low sensing cost, wide coverage, flexible deployment and strong computing capability, and enables a large number of applications in real-life, such as air quality monitoring [2], traffic information mapping [3], and feature description of interest points [4]. Task allocation [5] is an important part in MCS that can significantly impact the efficiency of MCS. It can be

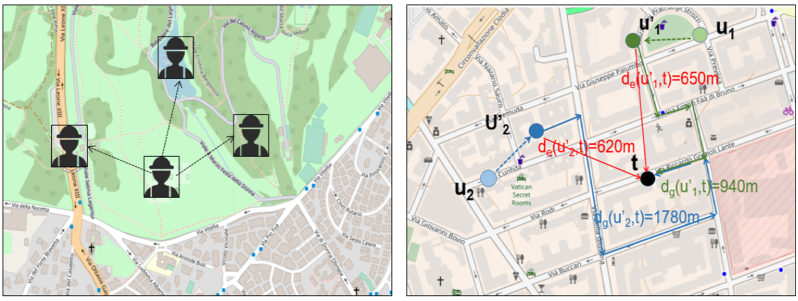
H. Gao and X. Wang (Eds.): CollaborateCom 2021, LNICST 406, pp. 20–37, 2021.
https://doi.org/10.1007/978-3-030-92635-9_2

distinguished into two scenarios: *Worker Selected Task* (WST) and *Server Allocated Task* (SAT). In this paper, we focus on the SAT.

In the SAT, workers need to upload their locations to MCS platforms for task allocation. When an adversary observes the location, he/she may infer worker's religion, home/working address, interest preference [6], etc. Therefore, it is necessary to protect worker's location privacy in task allocation.

In recent years, lots of work has been proposed for location privacy protection, and most of them focused on the obfuscation-based methods which allow workers to upload obfuscated locations instead of actual locations to MCS platforms. However, these solutions still have the following limitations.

(1) Most work on location protection and task allocation [7, 8] does not consider road networks. They assume that the location can be obfuscated to any point in a 2-dimensional distribution. As shown in Fig. 1(a), if the location is obfuscated to an unreasonable place, such as on a lake, in a forest, or on a railroad. An adversary can infer that this location is not true and use an attack model (such as Bayesian inference attack) to predict the worker's actual location, which may increase the risk of privacy leakage. Moreover, these methods assume that the distance between locations is measured by the Euclidean distance. When workers' mobility (such as driving) is restricted by road networks, this assumption may cause high utility loss. As shown in Fig. 1(b), the locations u_1 and u_2 are obfuscated to the locations $u_1^{'}$ and $u_2^{'}$. The Euclidean distance from $u_1^{'}$ to t and from $u_2^{'}$ to t on 2D are 650 m and 620 m, respectively. For optimal task allocation, task t should be assigned to worker u_2. However, the shortest distance from $u_1^{'}$ to t and from $u_2^{'}$ to t are 940 m and 1780 m (unavoidable detour) over road networks. The utility loss caused by $u_2^{'}$ reaches 1160 m, which is much higher than that of $u_1^{'}$ (290 m). Hence, the road network is a factor that cannot be ignored in privacy protection and task allocation.



(a) privacy (b) task allocation

Fig. 1. Influence of road networks on privacy and task allocation

(2) Existing work based on obfuscated methods [7, 9] usually takes the obfuscated locations provided by workers as the actual locations for task allocation. This means that they did not consider the influence of location obfuscation on task allocation efficiency, which is not reasonable and may cause high utility loss.

Recently, Geo-Graph-Indistinguishability (GeoGI) [10] has been presented in Location-Based Services (LBS) and an implementation method-Graph Exponential Mechanism (GEM) is proposed to solve the problem of privacy protection over road networks. However, task allocation is not considered in LBS. Therefore, GEM is not suitable to solve the problems raised in this paper. To tackle the above problems, we design a new privacy mechanism satisfying GeoGI and propose a Privacy Protection Task Allocation framework (PPTA) over road networks. In a nutshell, our contributions can be summarized as follows.

1) To protect workers' location privacy while preserving high utility in task allocation over road networks, we first propose to introduce Geo-Graph-Indistinguishability (GeoGI) to MCS and model a weighted directed graph according to the road network topology. Based on the graph, we formulate a linear programming to generate an optimal privacy mechanism, which can minimize the utility loss caused by location obfuscation under the constraints of GeoGI. We also improve the time-efficiency of the privacy mechanism generation by using δ-spanner graph.
2) To reduce the impact of location confusion on task allocation, we take the privacy mechanism proposed in PPTA as a key parameter to generate the task allocation scheme. With this idea, we formulate the problem of the optimal task allocation scheme generation as an integer programming, which aims to minimize workers' travel distance to task locations.
3) Experimental results on Roma taxi trajectory dataset show that PPTA can reduce the average travel distance of workers by up to 23.4% and increase privacy level by up to 21.5% compared to the existing differential privacy methods.

2 Preliminary

The MCS system consists of three parties, i.e., task requester, MCS platform and worker. The MCS platform is usually honest-but-curious, i.e., the platform assigns tasks to workers based on their obfuscated locations, but the platform is curious about the actual locations of workers. Therefore, we should protect workers' locations before uploading to the MCS platform. Consider the workers' mobility (such as driving) is restricted by road networks, we can represent the road network by a set of roads. When a road intersects, joins with other roads, or turns in a different direction, a connection is created. These connections divide roads into multiple road segments, which only connect with other road segments at their endpoints. Therefore, the road network of a city can be represented by a weighted directed graph $G = (V, E)$, where V is the vertex set representing road intersections and E is the edge set representing road segments. All vertices in V are on the road network, and for any pair of vertices $v_k, v_l \in V$, the weight of edge (v_k, v_l) is the shortest distance $d_g(v_k, v_l)$. Next, we introduce the privacy model and the adversary model.

2.1 Privacy Model

Geo-Graph-Indistinguishability (GeoGI) [10] is a novel location differential privacy model originally proposed for LBSs over road networks. It ensures that an adversary could not infer users' true locations from their released obfuscated locations. The location obfuscation process is to input an actual location u and output an obfuscated location u' through the probability matrix P. The probability matrix P is the key to realize GeoGI, which encodes the probability of obfuscating from a location to any location. GeoGI provides a feasible way to solve the problem of privacy protection for MCS over road networks.

The multiplicative distance between two distributions σ_1, σ_2 on some set $\mathcal{S}$ as $d_p(\sigma_1, \sigma_2) = sup_{S\in\mathcal{S}}|\ln\frac{\sigma_1(S)}{\sigma_2(S)}|$, with the convention that $|\ln\frac{\sigma_1(S)}{\sigma_2(S)}| = 0$ if both σ_1, σ_2 are zero and ∞ if only one of them is zero. Then, given $\varepsilon \in R^+$, Geo-Graph-Indistinguishability is defined as follows:

Definition 1. ***(Geo-Graph-Indistinguishability)*** [10]. *A probability matrix P on the road network $G = (V, E)$ satisfies Geo-Graph-Indistinguishability iff $\forall u_1, u_2, u'$ in V,*

$$d_p(P(u'|u_1), P(u'|u_2)) \leq \varepsilon d_g(u_1, u_2) \tag{1}$$

where $P(u'|u_1)$ is the probability of obfuscating u_1 to u', $d_g(u_1, u_2)$ is the shortest distance from u_1 to u_2 on the road network, and privacy budget ε is a parameter of GeoGI. The smaller ε, the higher privacy.

The definition can be also formulated as $\forall u_1, u_2, u' \in V$, $\frac{P(u'|u_1)}{P(u'|u_2)} \leq e^{\varepsilon d_g(u_1,u_2)}$. This formulation implies that GeoGI is an instance of d_x-privacy [11] proposed by Chatzikokolakis et al. The authors showed that an instance of d_x-privacy guaranteed strong privacy. Intuitively, this definition guarantees that if the obfuscation location is u', for any two locations u_1 and u_2 in V, the obfuscating probability of them to u' is approximate. Even though an adversary knows the obfuscated location u' and the probability matrix P, he/she cannot distinguish which one is the actual location. It is worth noting that GeoGI relies on a city's road network topology, and this results in the privacy protection level and utility varying depending on the road network even if the privacy parameter remains the same.

Specifically, in contrast to consider individual people's location [12] in LBS, task allocation efficiency in MCS depends on all the workers' locations. As shown in Fig. 2 (a), we assume a scenario with one worker u_1 and four tasks t_1, t_2, t_3, t_4. Due to task t_1, t_2 are closer to worker u_1, task t_1, t_2 will be allocated to worker u_1. When a new worker u_2 is added, as shown in Fig. 2(b), where $d_g(u_1, t_3) + d_g(u_2, t_2) < d_g(u_1, t_2) + d_g(u_2, t_3)$. For optimal task allocation, MCS platform will select worker u_1 perform task t_1, t_3 and worker u_2 perform task t_2, t_4. Therefore, the workers' location distribution must be considered in the generation of the probability matrix.

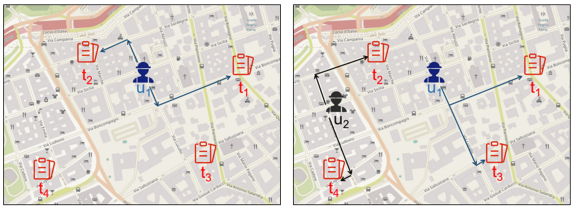


Fig. 2. Influence of workers distribution in task allocation

2.2 Adversary Model

We assume that workers' locations in MCS may suffer from Bayesian inference attack [6], i.e., if an adversary knows probability matrix P, worker's obfuscated location $u^{'}$ and location distribution $\pi(u)$, he/she can estimate the posterior distribution $\sigma(\widehat{u}|u')$ of the actual location u by resorting to the Bayes' Equation [13]. It is defined as follows:

$$\sigma(\widehat{u}|u') = \frac{P(u'|\widehat{u}) \cdot \pi(\widehat{u})}{\sum_{u* \in V} P(u'|u^*) \cdot \pi(u^*)} \tag{2}$$

where $u^{'}$ is the obfuscated location, $\widehat{u}$ is the location inferred by an adversary when observing the location $u^{'}$, u^* is any location in V, and $P\left(u^{'}|\widehat{u}\right)$ is the probability of obfuscating $\widehat{u}$ to u'.

We use the inference error (IE) proposed by Shokri et al. [14] to quantify the privacy level of a mechanism. The researchers translated location privacy into IE by measuring how accurately an adversary could infer the worker's actual location. Formally, IE can be formulated as follows:

$$IE\left(\pi_a, P, \sigma, d_g\right) = \sum_{u \in V} \sum_{u' \in V} \sum_{\widehat{u} \in V} \pi_a(u) P(u'|u) \sigma(\widehat{u}|u') d_g(u, u') \tag{3}$$

where $\pi_a(u)$ is the prior knowledge of workers' location distribution by an adversary, $P\left(u^{'}|u\right)$ is the probability of obfuscating u to $u^{'}$, $\sigma(\widehat{u}|u')$ is the posterior distribution of the actual location u, and $d_g(u, \widehat{u})$ is the shortest distance from u to $\widehat{u}$ on the road network.

Due to a probability matrix P satisfies GeoGI, it can limit the promotion of an adversary's posterior knowledge $\sigma(u|u')$ about workers' distribution over the prior knowledge π_a, i.e., $\sigma(u|u')/\pi_a \leq e^{\varepsilon D(R)}$ where $D(R)$ is the maximum distance of any two locations in region R. Please refer to [15] for the theoretical proof.

Consider an extreme situation, if an adversary knows the exact location of the worker through some ways in advance, then IE will always be zero. Therefore, we assume that the prior knowledge of the adversary $\pi_a(u)$ is equivalent to the public knowledge of workers' location distribution $\pi(u)$ (e.g., leaked by public check-ins [16]).

3 PPTA Framework

In this section, we introduce the PPTA framework. Figure 3 shows the overview of PPTA which consists of two modules: Location Obfuscation and Task Allocation Based on Obfuscated Locations. In the first module, the MCS platform needs to obtain workers' location distribution based on the historical sensing data and uses it as a parameter to generate a probability matrix via linear programming. After the platform generates the matrix, workers can download it into their smartphones, and then obfuscate their actual locations according to the probabilities encoded in the matrix. The obfuscated locations are uploaded to the platform for task allocation. In the second module, after receiving workers' obfuscated locations, the platform will assign tasks to proper workers, attempting to minimize the total traveling distance to the task locations. Since workers' uploaded locations are obfuscated, directly seeing them as actual locations for task allocation is not reasonable. Therefore, the probability matrix should be considered in the generation of a task allocation scheme for better allocation efficiency.

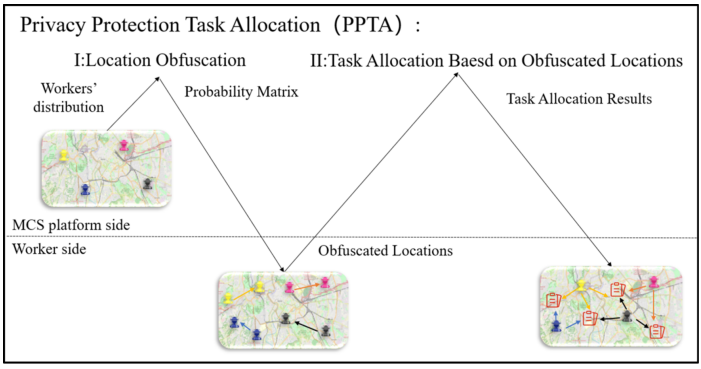


Fig. 3. PPTA framework

3.1 Location Obfuscation

In PPTA, each worker obfuscates his/her location through the probability matrix and uploads the obfuscated location to the MCS platform. Figure 4 shows an example of probability matrix, where workers' possible location is discrete into five vertices $\{v_1, v_2, v_3, v_4, v_5\}$. In this case, the probability matrix generated by the MCS platform is a 5×5 matrix. If the worker's actual location is v_2, according to the matrix, the probabilities that the worker outputs v_1, v_2, v_3, v_4, v_5 as the obfuscated location are 0.1, 0.3, 0.1, 0.1 and 0.4, respectively. Because the generation of probability matrix satisfies GeoGI, even if an adversary knows the obfuscated location and probability matrix, he/she cannot infer the actual location of the worker.

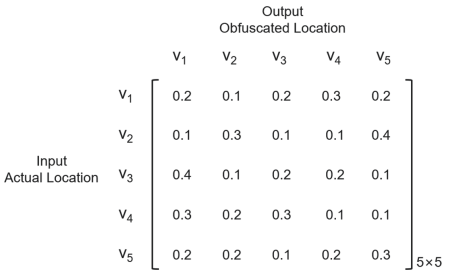


Fig. 4. Example of probability matrix

Although traditional privacy mechanisms provide a simple way to achieve privacy protection, these methods are independent of the prior knowledge of workers' location distribution which may cause high utility loss. To reduce this loss, we take the workers' distribution as a key parameter to generate the probability matrix P, which can be calculated from the workers' historical sensing data.

According to [17], the definition of utility loss (UL) is given:

$$UL = \sum_{v \in V} \sum_{v' \in V} \pi(v) P(v'|v) d_g(v, v') \tag{4}$$

where $\pi(v)$ is the workers' distribution, $P(v'|v)$ is the probability of obfuscating v to v', and $d_g(v, v')$ is the shortest distance from v to v' over road networks.

Given a group of locations $\{v_1, v_2, \ldots, v_n\}$ in region, a distance measure d_g, overall location distribution of workers $\pi(v)$ and privacy budget ε, the process of finding an optimal privacy mechanism P to minimize the utility loss is defined as follows:

Definition 2 [17]. *Given a prior location distribution π, a distance measure d_g and a privacy budget ε, a privacy mechanism P is GeoGI-OptUL (π, d_g) iff:*

1. *P is geo-graph-indistinguishability and*
2. *for all mechanisms P', if P' is geo-graph-indistinguishability then $UL(P, \pi, d_g) \leq UL(P', \pi, d_g)$.*

Note that GeoGI-OptUL (π, d_g) optimizes UL given a privacy constraint of GeoGI. Now, we can formulate the problem of privacy mechanism generation (PMG) as a linear programming, which can minimize the utility loss:

$$\min_{P} \sum_{v \in V} \sum_{v' \in V} \pi(v) P(v'|v) d_g(v, v') \tag{5}$$

$$subject\ to: P(v'|v_1) \leq e^{\varepsilon d_g(v_1, v_2)} P(v'|v_2), \quad v_1, v_2, v' \in V \tag{6}$$

$$\sum_{v' \in V} P(v'|v) = 1, \quad v \in V \tag{7}$$

$$P(v'|v) \geq 0, \quad v, v' \in V \tag{8}$$

where constraint (6) is the definition of GeoGI, the shortest distance d_g from workers to tasks is calculated by Dijkstra's algorithm on the road network and updated by the MCS platform before each round of task allocation. Therefore, the road network is considered in PMG. Constraints (7–8) are the basic requirements of probability. Obviously, PMG satisfies GeoGI.

It should be noted that PPTA does not need workers to upload locations frequently, only once in PPTA. The location obfuscation runs completely in the worker's smartphone, so no one else knows the worker's actual location.

3.2 Task Allocation Based on Obfuscated Locations

We consider a scenario where M tasks need to be allocated to N workers ($N < M$). After the locations of N workers have been obfuscated and uploaded to the MCS platform, the task allocation scheme designs by the MCS platform can be represented by an indicator matrix $X = \{x_{i,j}\}_{N \times M}$, where the matrix element $x_{i,j}$ indicates whether task j is allocated to worker i, i.e., $x_{i,j}$= 1 if task j is assigned to worker i; otherwise, $x_{i,j}$ = 0. The matrix X needs to satisfy the constraints: (1) $\sum_i x_{i,j} \geq p, \forall j (j = 1, \ldots, M)$, i.e., at least p workers are required to perform a task. (2) $\sum_j x_{i,j} \geq q, \forall i (i = 1, \ldots, N)$, i.e., each worker needs to perform at least q tasks. For the MCS platform, considering the phenomenon of malicious uploading data by workers, constraint (1) is to improve the quality of data. For workers, constraint (2) is to accept more tasks in a single task allocation without uploading location multiple times can not only improve the revenue, but also reduce the risk of privacy leakage.

The goal of task allocation is to ensure each task should be assigned to workers and the total travel distance of workers is minimized, which is also known as a common metric to measure the task allocation efficiency. Hence, we formulate the optimal task allocation scheme as follows, and solve it with an integer programming.

$$\min_x \sum_{u' \in V} \sum_{t \in V} d_g'(u', t) x(u', t) \tag{8}$$

$$subject\ to: x(u', t) \in \{0, 1\}, \forall u', t \in V \tag{9}$$

$$\sum_{t \in V} x(u', t) \geq q, \quad \forall u' \in V \tag{10}$$

$$\sum_{u' \in V} x(u', t) \geq p, \quad \forall t \in V \tag{11}$$

where $d_g'(u', t)$ represents the travel distance from the obfuscated location u' to the task t over road networks. x is the task allocation scheme generated by the MCS platform according to the obfuscated locations.

Due to the MCS platform receives the obfuscated locations, it is unreasonable to directly regard them as the actual locations for task allocation. Although the privacy

mechanism designed in this paper has taken it into account, it still brings high utility loss. According to the definition of probability matrix P, any two locations are obfuscated to the same location is approximate. That means, for an obfuscated location, all locations in region may be its actual location. Therefore, we combine the probability matrix into the generation of task allocation scheme and use the distance from all possible locations to the task instead of the distance from the obfuscated location to the task. The mathematical relationship is as follows:

$$d_g^{'}(u', t) = \frac{\sum_{u \in V} \pi(u) P(u'|u) d_g(u, t)}{\sum_{u \in V} \pi(u) P(u'|u)} \tag{12}$$

From the above formulation, we can find that the higher the obfuscated probability of a location, the more likely it is to be the actual location, and the larger its distance weight, the lower the effect of location obfuscation on task allocation. Finally, by plugging Eq. (12) into Eq. (8), we can get a final objective function:

$$\min_{x} \sum_{u' \in V} \sum_{t \in V} \frac{\sum_{u \in V} \pi(u) P(u'|u) d_g(u, t)}{\sum_{u \in V} \pi(u) P(u'|u)} x(u', t) \tag{13}$$

The value range of the variable x in the model is limited to integer, thus this is an integer programming. The above two mathematical models can be solved by standard LP approaches, such as the simplex methods, or the advanced program solver (e.g., *CPLEX, Lingo*).

3.3 Speed-Up with δ-Spanner Graph

In the process of PMG, the number of constraints (6) is $O(|V|^3)$, which makes the method proposed in this paper difficult to extend to large-scale regions in real-life. Considering the total number of constraints to generate the task allocation scheme is $O(|N||M|)$, which is far less than $O(|V|^3)$. Thus, we only need to optimize PMG, which is the most time-consuming part in PPTA. Some common approaches, such as the dual form of linear programming, can be used to speed up PMG. However, although there are fewer constraints, the constraints in the dual problem will become more complex, so it is not practical.

We speed up PPTA by using δ-spanner graph. It ensures that for a given obfuscated location, it compares whether the obfuscating probability of adjacent locations in region satisfies GeoGI, rather than any two locations. According to [17], we construct a δ-spanner graph, which contains all the vertices in a weighted directed graph but reduces the number of edges. Stretch factor δ is an important parameter of δ-spanner graph, which represents the maximum ratio of the distance between any two vertices in two graphs. The definition is as follows:

Definition 3. ***(Dilation)*** [17]. *Let $G_\delta = V$, E_δ be a spanner graph. The dilation of G_δ is calculated as:*

$$\delta = \max_{v \in V, v' \in V, v \neq v'} \frac{d_{g_\delta}(v, v')}{d_g(v, v')} \tag{14}$$

A spanner with dilation δ is called a δ-spanner graph.

Now, to speed up PPTA, we introduce δ-spanner graph to GeoGI and the following theorem holds:

Theorem 1 [18]. *If $G_\delta(V, E_\delta)$ is a δ-spanner graph, and a probability matrix P satisfies:*

$$P(v'|v_1) \leq e^{\frac{\varepsilon}{\delta} d_{g_\delta}(v_1, v_2)} P(v'|v_2), \quad (v_1, v_2) \in E_\delta, v' \in V \tag{15}$$

Then, P satisfies Geo-Graph-Indistinguishability.

Proof. According to Eq. (14), we can obtain

$$d_{g_\delta}(v, v')/\delta \leq d_g(v, v'), \quad \forall v, v' \in V \tag{16}$$

By using Eq. (6) and Eq. (16), we can derive

$$P(v'|v_1) \leq e^{\frac{\varepsilon}{\delta} d_{g_\delta}(v_1, v_2)} P(v'|v_2) \leq e^{\varepsilon d_g(v_1, v_2)} P(v'|v_2), \ \forall (v_1, v_2) \in E_\delta, v' \in V \tag{17}$$

This concludes the proof.

According to Definition 3 and Theorem 1, the mathematical model of PMG is updated:

$$\min_P \sum_{v \in V} \sum_{v' \in V} \pi(v) P(v'|v) d_{g_\delta}(v, v') \tag{18}$$

$$subject\ to: P(v'|v_1) \leq e^{\frac{\varepsilon}{\delta} d_{g_\delta}(v_1, v_2)} P(v'|v_2), \quad \forall (v_1, v_2) \in E_\delta, v' \in V \tag{19}$$

$$\sum_{v' \in V} P(v'|v) = 1, \ v \in V \tag{20}$$

$$P(v'|v) \geq 0, \ v, v' \in V \tag{21}$$

The number of constraints (19) is $O(|E_\delta||V|)$ by using δ-spanner graph. For a δ-spanner graph [17], $|E_\delta| = \frac{|V|}{\delta - 1}$, thus the number of constraints in PMG can be reduced from $O(|V|^3)$ to $O(|V|^2)$ approximately. Following previous work [17], when δ equal to 1.08, the experimental effect is the best.

4 Evaluation

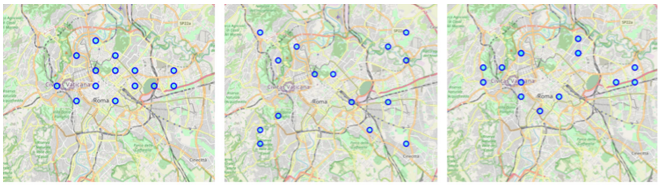
In this section, we first evaluate the performance of the proposed PPTA framework in terms of privacy and utility with a real-world dataset. Then, we evaluate the time-efficiency of PMG before and after the optimization of δ-spanner graph.

4.1 Experiment Configurations

Evaluation Scenario

We conduct experiments by using a publicly real-world taxi trajectory dataset in Roma [23]. The dataset contains GPS coordinates of approximately 320 taxis collected over 30 days and some of them are selected for the experiments. The longitude range of the selected dataset is (12.418, 12.574) and the latitude range is (41.859, 41.947). Most of the data is in the central of Roma and a small percentage in the suburbs. We select to use a taxi dataset since taxi services can be regarded as a MCS application type (taxi driver can be considered as a worker, passenger can be considered as a task).

To evaluate the PPTA, we set up the experiments with different parameters. The privacy budget ε ranges from ln (2) to ln (8). ε is usually chosen by workers, for simplicity, we set the same ε for each worker in the experiment. In each round of task allocation (a round is set to 1 h), the number of workers (N) ranges from 15 to 40, and the number of tasks (M) ranges from 45 to 120. Note that before each round of task allocation, we learn workers' location distribution $\pi(u)$ according to workers' historical sensing data. We also conduct experiments for different task distributions (Fig. 5), which are compact, scattered and hybrid. Finally, we change the size of the region to evaluate the time-efficiency of PMG before and after the optimization of δ-spanner graph.



(a) compact (b) scattered (c) hybrid

Fig. 5. Three types of task distribution

In this paper, we use Lingo to solve two linear programming problems, and python is used for experiments. All experiments are conducted on inter (R) core (TM) i7-4710 hq CPU@2.5 GHz, 8 GB RAM, win10 OS.

Evaluation Metrics

ATD (Average Travel Distance). Lots of work [7, 9] considers workers' travel distance to task locations as an important factor in task allocation. Following previous work, we use ATD as the utility metric, which can be calculated as:

$$ATD = \sum_{(u,t)\in X} d_g(u,t)/|x| \tag{22}$$

where $|x|$ is the total number of tasks allocated to workers. The smaller the ATD, the lower the utility loss is.

IE (Inference Error). An adversary can infer worker's actual location by resorting to the Bayes' Equation, if he/she knows the worker's upload location, workers' location distribution and probability matrix. We use IE (Eq. (3)), i.e., the expected distortion from the inferenced location (by adversary) to the actual location, to quantify the privacy level of PPTA. If IE is smaller, the privacy level of the mechanism is lower.

Baselines

Laplace Mechanism (LAP) [7]. Laplace is a traditional differential privacy mechanism and tends to obfuscate a location to its nearby location with high probability. Its probability distribution is derived from a two-dimensional version of the Laplace distribution as follows.

$$P_{lap}(u'|u) \propto e^{-\varepsilon \frac{d_e(u,u')}{D(R)}} \tag{23}$$

where d_e is the Euclidean distance. $D(R)$ is the maximum distance between any two locations in the region R.

Exponential Mechanism (EXP) [19]. Exponential mechanism is also widely used to achieve differential privacy. In the design of Exponential mechanism, a scoring function needs to be modeled to obtain high utility. Given a location, a better obfuscation should be assigned a higher score. In MCS, a higher score is preferred for the location obfuscation that leads to lower utility loss. With this idea, we design the following Exponential mechanism.

$$P_{\exp}(u'|u) \propto e^{\frac{\varepsilon}{2} \cdot \left(1 - \frac{d_g(u,u')}{\max_{u* \in R} d_g(u,u*)}\right)} \tag{24}$$

Graph Exponential Mechanism (GEM) [10]. Graph exponential mechanism employs the idea of exponential mechanism. This mechanism considers road networks so that high utility can be expected. Moreover, since this mechanism satisfies GeoGI, strong privacy based on differential privacy is guaranteed. The obfuscated probability is as follows.

$$P_{gem}(u'|u) \propto e^{-\frac{\varepsilon}{2} d_g(u,u')} \tag{25}$$

No Privacy Protection. No privacy protection means that the MCS platform knows all workers' actual locations, which can be regarded as the lower bound of ATD for task allocation based on obfuscated locations.

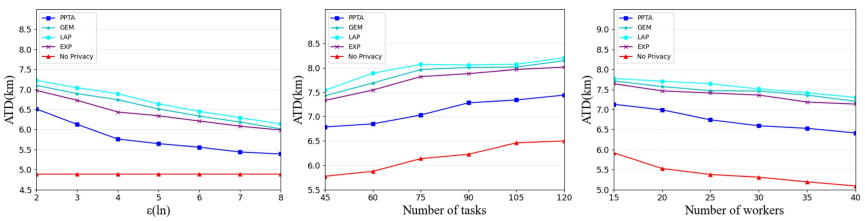
4.2 Experimental Results

Utility

In the experiment, we first evaluate PPTA in term of utility (task allocation efficiency). If ATD is closer to No Privacy, the utility loss is lower. It can be seen from Fig. 6(a) that all the methods (except No Privacy) will lead to the continuous decrease of ATD, as the privacy budget ε increases. According to the definition of differential privacy, a

larger ε denotes the lower privacy level, which leads to a location will be obfuscated to its nearby location with a high probability. Thus, ATD shows a downward trend. In addition, privacy budget ε cannot influence on the No Privacy, so the ATD of No Privacy stays the same. In comparison of the four privacy mechanisms, the ATD generated by Laplace mechanism is the largest. This is because Laplace mechanism does not consider the characteristics of road networks and use Euclidean distance, resulting in traditional differential privacy mechanisms are not suitable to provide location protection over road networks. Graph exponential mechanism and Exponential mechanism add road network constraints in the design process, so their utility loss is lower than Laplace mechanism. For the Exponential mechanism, we redesign its scoring function so that a location will be obfuscated with a higher probability to a location with smaller utility loss. As for Graph exponential mechanism, it is a privacy mechanism proposed in LBS, which does not consider task allocation. Therefore, compared with the Exponential mechanism, the application of Graph exponential mechanism to MCS will result in a larger utility loss. From the experimental results, we can find that these methods still have a certain gap with PPTA in term of utility. The main reason for this situation is that PPTA considers the influence of location obfuscation on task allocation and reduces the influence in the optimization model.

Figure 6(b) describes the relationship between the number of tasks and ATD. It can be seen from Fig. 6(b) that with the increase of number of tasks, ATD shows an upward trend. This is because when the number of tasks increases, each worker needs to perform more tasks to complete the task requirements of each round. From the experimental results, we can find that ATD generated by PPTA is always smaller than the other three privacy mechanisms. Figure 6(c) describes the relationship between the number of workers and ATD. As the number of workers increases, ATD shows a downward trend. This is because workers who upload their locations will be assigned tasks (otherwise privacy will be sacrificed in vain) and there are more choices to assign tasks for the platform. Compare with other privacy mechanisms, PPTA generates the smallest ATD regardless of the number of workers or tasks and has a stable performance.



(a) N=15, M=45, hybrid (b) ε=ln (4), N=15, hybrid (c) ε=ln (4), M=120, hybrid

Fig. 6. Varying privacy budget, task number, and worker number

We also evaluate the effect of different task distributions on utility: compact distribution, scattered distribution, and hybrid distribution. From previous experiments, we can find that Exponential mechanism performed better than Laplace mechanism and

Graph exponential mechanism in term of utility, so we only compare Exponential mechanism and PPTA. As shown in Fig. 7, we can find that the scattered distribution generates the largest ATD of three task distributions, and the compact distribution has the smallest ATD. This is because the distance between tasks under the scattered distribution is longer than other two task distributions, and workers need to travel a longer distance to accomplish all tasks. From the experimental results, we can also find that no matter compact, scattered or hybrid, PPTA generates much smaller ATD than Exponential mechanism. This means that our proposed method can achieve stable performance across different distributions. Noted that compared with the scattered and hybrid distribution, ATD generated by PPTA and Exponential mechanism under the compact distribution is the closest to no privacy, which indicates that the compact distribution has the lowest influence on utility. On the contrary, the scattered distribution has the greatest influence on utility.

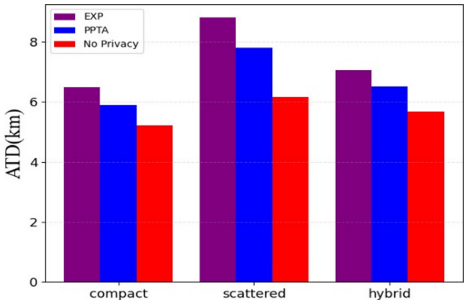


Fig. 7. Varying task distribution (ε = ln (4), N = 15, M = 45)

Privacy

We next evaluate PPTA in term of privacy. If IE is smaller, the privacy level of mechanism is lower. As shown in Fig. 8, as ε increases, all methods lead to the continuous decrease of IE. According to Eq. (3), we can find that a larger ε means the adversary's inferenced location is closer to the actual location. Thus, IE shows a downward trend. With the comparison of other three mechanisms (LAP, GEM, EXP), the IE generated by PPTA is the largest which indicates PPTA provides the best privacy protection under the same privacy budget ε. From the experimental results, when ε is ln (5), the largest difference in IE between PPTA and other privacy mechanisms. It means that when consider both privacy and utility in MCS, the privacy budget ε set as ln (5) has a best effect. As shown in Fig. 6(a) and Fig. 8, when the privacy budget ε is constantly increasing, ATD of PPTA is closer to No Privacy and the task allocation efficiency is improved. However, IE is constantly decreasing and the privacy level will be lower.

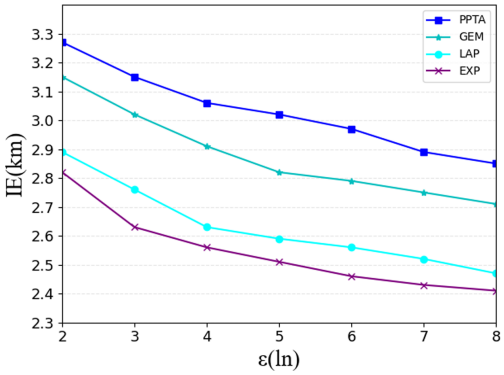


Fig. 8. Varying privacy budget (N = 15, M = 45, hybrid)

Time-Efficiency

In Sect. 3.3, we have theoretically proved that the number of constraints in PMG is reduced to $O(|V|^2)$ by using δ-spanner graph, where $|V|$ denotes the number of vertices in the graph. We repeat the experiments many times and record the mean time, which takes about 1 min 45 s (19 s after the optimization) and 3 s to generate probability matrix and task allocation scheme, which is totally acceptable in real-life MCS applications.

As shown in Fig. 9(a) (b), after optimization, the number of iterations and the corresponding computation time of PMG have been significantly reduced when the location number equals 25, 49, 81. From the experimental results, we can find that the larger the location number, the more significant the optimization effect. When the locations' number reaches 81, the computation time is shortened by five times and the number of iterations is reduced by nearly half after optimization.

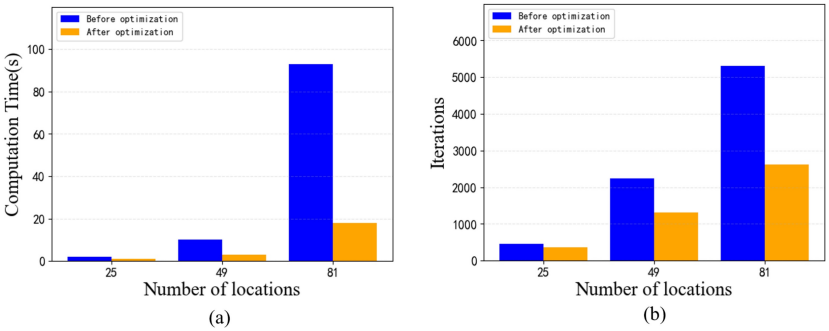


Fig. 9. Varying region size (ε = ln (4), N = 15, M = 45, hybrid)

5 Related Work

In recent years, a variety of location privacy protection approaches have been proposed, such as anonymity [20], encryption [21], etc. However, these works have the drawback of dependence on trustful platforms and high cost. To address these problems, we focus on obfuscation-based methods, and differential privacy has been applied to address location privacy issues in MCS.

In the context of sparse MCS, Wang et al. [18, 22] propose a privacy protection framework, which takes into account the level of privacy, the prior knowledge about workers' location distribution, and the data quality loss due to location obfuscation. Particularly, the framework can provide a guaranteed level of differential and distortion privacy with reduced data quality loss in Sparse MCS applications. However, they only evaluate the data quality loss and do not consider the privacy level of the proposed framework. Yang et al. [7] analyzes the shortcomings of existing works and propose a mixed integer nonlinear programming model which aims at minimizing workers' travel distance. It uses differential geo-obfuscation to protect workers' location privacy regardless of adversaries' prior knowledge, without the involvement of any trustful third-party. However, it does not consider the adversary attack and use Euclidean distance as a metric, which still has some defects in privacy and utility. Liu et al. [8] consider two kinds of task allocation scenarios, multi-task with few workers and few tasks with multi workers. They design the multi-objective optimization model and propose the W-ILP and C-ILP algorithms to select workers with the minimum total incentive payments and minimum total travel distance. However, they did not consider the road networks, which may cause insufficiencies in terms of privacy and utility.

6 Conclusion

In this paper, we have proposed a Privacy Protection Task Allocation framework (PPTA) to protect workers' location privacy for MCS over road networks. We introduce GeoGI and model a weighted directed graph according to the road network topology. Then, we formulate a linear programming to generate an optimal privacy mechanism. It considers the level of privacy protection, the prior knowledge about workers' location distribution and the utility loss due to location obfuscation. We also improve the time-efficiency of the privacy mechanism generation by using δ-spanner graph. Finally, we design an optimal task allocation scheme based on obfuscated locations by using an integer programming, which aims to minimize workers' travel distance to task locations. Experimental results on Roma taxi trajectory dataset show that PPTA can reduce average travel distance of workers by up to 23.4% and increase privacy level by up to 21.5% compared to the existing differential privacy methods.

In this paper, we assume that the workers are not related to each other. However, this assumption may not be reasonable. Adversary may learn about workers' correlation from their check-in records, which may cause unexcepted privacy leakage. In the future, we will aim at researching this correlation attack and will propose a corresponding solution.

Acknowledgments. The authors would also like to appreciate the anonymous reviewers for their valuable suggestions, which lead to a substantial improvement of this paper. This research has been funded by the National Natural Science Foundation of China (Grant No. 61672468, 61702148).

References

1. Capponi, A., Fiandrino, C., Kantarci, B., Foschini, L., Kliazovich, D., Bouvry, P.: A survey on mobile crowdsensing systems: challenges, solutions, and opportunities. IEEE Commun. Surv. Tutorials **21**(3), 2419–2465 (2019)
2. Fiandrino, C., et al.: CrowdSenSim: a simulation platform for mobile crowdsensing in realistic urban environments. IEEE Access **5**, 3490–3503 (2017)
3. Tong, Y., Chen, L., Shahabi, C.: Spatial crowdsourcing: challenges, techniques, and applications. Proc. VLDB Endow. **10**(12), 1988–1991 (2017)
4. Chon, Y., Lane, N.D., Li, F., Cha, H., Zhao, F.: Automatically characterizing places with opportunistic crowdsensing using smartphones. In: Proceedings of the 2012 ACM Conference on Ubiquitous Computing, New York, USA, pp. 481–490 (2012)
5. Wang, J., Wang, L., Wang, Y., Zhang, D., Kong, L.: Task allocation in mobile crowd sensing: state-of-the-art and future opportunities. IEEE Internet Things J. **5**(5), 3747–3757 (2018)
6. Chen, J., Ma, H., Zhao, D., Liu, L.: "Correlated differential privacy protection for mobile crowdsensing. IEEE Trans. Big Data 1 (2017)
7. Wang, L., Yang, D., Han, X., Wang, T., Zhang, D., Ma, X.: Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation. In: Proceedings of the 26th International Conference on World Wide Web, Perth Australia, pp. 627–636 (2017)
8. Liu, Y., Guo, B., Wang, Y., Wu, W., Yu, Z., Zhang, D.: TaskMe: multi-task allocation in mobile crowd sensing. In: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, New York, USA, pp. 403–414 (2016)
9. Qiu, C., Squicciarini, A., Li, Z., Pang, C., Yan, L.: Time-efficient geo-obfuscation to protect worker location privacy over road networks in spatial Crowdsourcing. In: CIKM 20: Proceedings of the 29th ACM International Conference on Information & Knowledge Management. CIKM (2020)
10. Takagi, S., Cao, Y., Asano, Y., Yoshikawa, M.: Geo-Graph-indistinguishability: protecting location privacy for LBS over road networks. In: Foley, S.N. (ed.) DBSec 2019. LNCS, vol. 11559, pp. 143–163. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22479-0_8
11. Chatzikokolakis, K., Andrés, M.E., Bordenabe, N.E., Palamidessi, C.: Broadening the scope of differential privacy using metrics. In: Privacy Enhancing Technologies, Berlin, Heidelberg, pp. 82–102 (2013)
12. Qiu, C., Squicciarini, A.C., Pang, C., Wang, N., Wu, B.: Location privacy protection in vehicle-based spatial Crowdsourcing via geo-indistinguishability. IEEE Trans. Mobile Comput. **99**, 1 (2020)
13. Yang, B., Sato, I., Nakagawa, H.: Bayesian differential privacy on correlated data. In: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data - SIGMOD '15, Melbourne, Victoria, Australia, pp. 747–762 (2015)
14. Shokri, R., Theodorakopoulos, G., Boudec, J.L., Hubaux, J.: Quantifying location privacy. In: 2011 IEEE Symposium on Security and Privacy, pp. 247–262 (2011)

15. Andrés, M.E., Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C.: Geo-indistinguishability: differential privacy for location-based systems. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security, New York, USA, pp. 901–914 (2013)
16. Dingqi, Y., et al.: Participatory Cultural Mapping Based on Collective Behavior Data in Location-Based Social Networks. https://dl.acm.org/doi/abs/10.1145/2814575. Accessed 21 July 2021
17. Bordenabe, N.E., Chatzikokolakis, K., Palamidessi, C.: Optimal geo-indistinguishable mechanisms for location privacy. In: Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, New York, USA, pp. 251–262 (2014)
18. Wang, L., Zhang, D., Yang, D., Lim, B.Y., Han, X., Ma, X.: Sparse mobile crowdsensing with differential and distortion location privacy. IEEE Trans. Inform. Forensic Secur. **15**, 2735–2749 (2020)
19. McSherry, F., Talwar, K.: Mechanism design via differential privacy. In: 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07), pp. 94–103 (2007)
20. Duckham, M., Kulik, L.: A formal model of obfuscation and negotiation for location privacy. In: Pervasive Computing, Berlin, Heidelberg, pp. 152–170 (2005)
21. Xiong, J., et al.: A personalized privacy protection framework for mobile crowdsensing in IIoT. IEEE Trans. Industr. Inf. **16**, 4231–4241 (2020)
22. Wang, L., Zhang, D., Yang, D., Lim, B.Y., Ma, X.: Differential location privacy for sparse mobile crowdsensing. In: 2016 IEEE 16th International Conference on Data Mining (ICDM), Barcelona, Spain, pp. 1257–1262 (2016)
23. CRAWDAD dataset. https://crawdad.org/roma/taxi/20140717. Accessed July 2014