Personalized Privacy-preserving Task Allocation for Mobile Crowdsensing

Zhibo Wang, *Senior Member, IEEE,* Jiahui Hu, Ruizhao Lv, Jian Wei, Qian Wang, *Member, IEEE,* Dejun Yang, *Member, IEEE,* and Hairong Qi, *Fellow, IEEE*

Abstract—Location information of workers are usually required for optimal task allocation in mobile crowdsensing, which however raises severe concerns of location privacy leakage. Although many approaches have been proposed to protect the locations of users, the location protection for task allocation in mobile crowdsensing has not been well explored. In addition, to the best of our knowledge, none of existing privacy-preserving task allocation mechanisms can provide personalized location protection considering different protection demands of workers. In this paper, we propose a personalized privacy-preserving task allocation framework for mobile crowdsensing that can allocate tasks effectively while providing personalized location privacy protection. The basic idea is that each worker uploads the obfuscated distances and personal privacy level to the server instead of its true locations or distances to tasks. In particular, we propose a Probabilistic Winner Selection Mechanism (PWSM) to minimize the total travel distance with the obfuscated information from workers, by allocating each task to the worker who has the largest probability of being closest to it. Moreover, we propose a Vickrey Payment Determination Mechanism (VPDM) to determine the appropriate payment to each winner by considering its movement cost and privacy level, which satisfies the truthfulness, profitability and probabilistic individual rationality. Extensive experiments on the real-world datasets demonstrate the effectiveness of the proposed mechanisms.

Index Terms—Mobile Crowdsensing; Task Allocation; Differential Privacy, Personalized Privacy-preserving

1 Introduction

Nowadays, the ubiquity of mobile devices equipped with various functional built-in sensors (e.g., camera, microphone, accelerometer, GPS) and the increasingly powerful wireless network have enabled the prosperity of mobile crowdsensing. The new computing paradigm that leverages the power of the crowd supports mobile users to opportunistically perform tasks according to their interests and schedule. Specifically, the mobile crowdsensing systems mainly collect spatio-temporal data from environments, social and others. A typical mobile crowdsensing system consists of data requesters, a server and mobile users (workers), where the server publishes spatio-temporal tasks outsourced by data requesters to mobile users, and then mobile users use their mobile devices to complete the published tasks and upload the collected data to the server. Mobile crowdsensing has a wide range applications in environmental sensing [1], journalism [2], crisis response [3] and urban planning [4]. As a representative, the commercial app Waze [5], is a popular traffic monitoring and route navigation system that collects real-time traffic data from mobile users in a crowdsourcing way.

Task allocation is one of the most important problem in mobile crowdsensing, which relies on the distances between tasks and mobile users to assign tasks appropriately. However, the location information may be disclosed during the task allocation process, especially when the server cannot be trusted (e.g., the server may be benefited by selling the location information to third parties). Adversaries who obtain the locations can stage extensive attacks such as physical surveillance, stalking, identity theft, and breach of sensitive information (e.g., health status, political and religious views). Hence, disclosing true locations to the server may be harmful to workers, which further discourages them from engaging in crowdsensing. This problem is more severe for the mobile users who are not allocated with any task at all, since their locations are disclosed to the server without receiving any payoff. Therefore, location privacy should be carefully considered in task allocation of mobile crowdsensing systems.

Existing works on mobile crowdsensing mainly focus on maximizing utility of task allocation and designing incentive mechanisms to improve user participation, while privacy protection has not been extensively explored. Recently, several works begin to address the problem of task allocation in mobile crowdsensing with location privacy protection. Liu et al. [6] applied the economic model for location privacy preserving and proposed a mechanism to leave out the bids and tasks assignment processes which have a risk of privacy leakage. However, this mechanism can only protect the location privacy against the eavesdroppers and hackers over the communication channel while the malicious sever was out of its scope. Spatial cloaking is the most straightforward technique that blurs a user's exact location into a spatial region in order to preserve the

- *Zhibo Wang, Jiahui Hu, Ruizhao Lv, Jian Wei and Qian Wang are with Key Laboratory of Aerospace Information Security and Trusted Computing, Ministry of Education, School of Cyber Science and Engineering, Wuhan University, 430072, P.R. China. E-mail: {zbwang, jiahuihu, ruizhaolv, wj9528, qianwang}@whu.edu.cn*
- *Dejun Yang is with the Department of Computer Science at Colorado School of Mines, CO, USA. E-mail: djyang@mines.edu*
- *Hairong Qi is with the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville USA. E-mail: hqi@utk.edu*

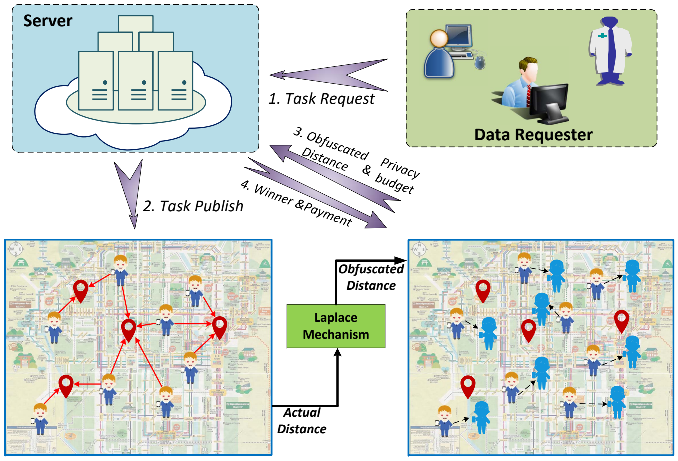


Fig. 1. The proposed framework of personalized privacy-preserving task allocation in mobile crowdsensing with obfuscated distance.

location privacy [7]–[9]. However, the privacy guarantees can be easily downgraded if adversaries hold certain prior knowledge, which is known as inference attack. In contrast, the authors in [10], [11] introduced differential privacy into task allocation in mobile crowdsensing, providing theoretically guaranteed location privacy protection regardless of adversaries' prior knowledge. However, To et al. [10] assumed that there is a trusted third party (cellular service providers) to play the coordination role between the server and workers, while in practice the cellular service providers have no motivation to participate. As for [11], the proposed geo-obfuscation function is related to the task locations and need to be updated if the task distribution is transformed, which may discourage the workers who need to download the function before obfuscating their actual locations. Most importantly, both of them employ the same level of privacy protection for all workers, which cannot satisfy the different privacy demands of workers. Consequently, some workers may get insufficient privacy protection, while others get over-protected.

In this paper, we propose a personalized privacy-preserving task allocation framework for mobile crowdsensing that can allocate tasks effectively while providing personalized location privacy protection. As shown in Figure 1, the basic idea is that, each worker utilizes Laplace Mechanism [12] to obfuscate the distances between itself to tasks with its personal privacy protection level, and uploads the obfuscated distances as well as its personal privacy protection level to the server, instead of uploading its true location or true distances to tasks. The server further selects the winner for each task and determines the appropriate payment for the winners. In particular, we propose a Probabilistic Winner Selection Mechanism (PWSM) to minimize the total travel distance with the obfuscated information from workers, by allocating each task to the worker who has the largest probability of being closest to it and eliminating the winner conflict problem. Moreover, we propose a Vickrey Payment Determination Mechanism (VPDM) to determine the appropriate payment to each winner by considering its movement cost and privacy leakage, which satisfies the truthfulness, profitability and probabilistic individual rationality.

The contributions of this paper are summarized as follows.

- To the best of our knowledge, this is the first work that realizes personalized location privacy protection for task allocation in mobile crowdsensing, which enables users to balance the payoff and privacy leakage, and encourages users to participate in mobile crowdsensing with theoretical guarantee of location privacy preservation.
- We propose a Probability Compare Function (PCF) to determine which worker has a higher probability of being closer to a task for two workers. Based on the PCF, we further propose the PWSM to minimize the total travel distance with the obfuscated information and personal privacy levels from workers.
- We propose the VPDM to determine the appropriate payment to each winner by considering its movement cost and privacy level, which satisfies the truthfulness, profitability and probabilistic individual rationality.
- We conduct extensive experiments on the real-world check-ins datasets and compare the proposed PWSM with two benchmarks. The experimental results demonstrate the effectiveness of the proposed mechanisms.

The remainder of this paper is organized as follows. We discuss the related works in Section 2. We present a high-level overview of our framework and formulate the problems in Section 3. The proposed winner selection mechanism and payment determination mechanism are presented in Section 4 and 5, respectively. We evaluate the performance of the proposed mechanisms in Section 7 and finally conclude the paper in Section 8.

2 Related Work

We briefly discuss the related work from the following aspects: task allocation, location privacy and personalized privacy.

2.1 Task Allocation

Spatial tasks require the workers to be at a specific place in order to fulfill a task. The locations of the workers play an important role in task allocation as the workers need to travel to locations of interested tasks to perform sensing jobs. There have been several works on minimizing the travel distance in task allocation. Guo et al. [13] proposed a framework for optimizing tasks allocation considering the time-sensitive tasks and delay-tolerant tasks, respectively. In [14], the authors proposed a task allocation framework for multi-task environments with the objective of minimizing the travel distance. Moreover, there are also some works focusing on the sensing data quality, rewards, budget and social costs. In [15], the authors proposed a novel framework called CCS-TA to dynamically select a minimum number of sub-areas for sensing task allocation in each sensing cycle while guaranteeing the quality of sensing data. In [16], the authors considered the sparse crowdsensing and proposed a task allocation framework with the objective of maximizing the quality of sensing data. He et al. [17] considered sensing tasks with different requirements of quality and proposed a mechanism which allocates tasks to mobile users who are constrained by time budgets with the objective of maximizing total rewards of platform. Wang et al. [18] considered the time-sensitive and location-dependent crowdsensing systems with random arrivals and proposed a task allocation mechanism for maximizing profits of participants. In [19] [20], the authors considered the piggyback crowdsensing and proposed to allocate tasks for maximizing the coverage quality of the sensing task while satisfying the incentive budget constraints. A demand-based dynamic task allocation mechanism was proposed in [21] to balance the participation among location-dependent tasks. Lin et al. [22] considered the bid privacy of users and proposed two privacy-preserving task allocation frameworks which protect the privacy of bids while achieving approximate social cost minimization.

2.2 Location Privacy

Originating from the spatio-temporal privacy mechanisms in the location based service (LBS), the spatial cloaking technique can also be used for privacy-preserving task allocation. Spatial cloaking hides the worker's location inside a cloaked region so that the adversary cannot attain the actual location of the worker [7]–[9]. In practice, there are many mobile crowdsensing applications that do not require exact locations, such as air quality monitoring. However, he privacy guarantee can be easily downgraded if adversaries hold certain prior knowledge. Regardless of the adversaries' prior knowledge, differential privacy [23] was applied for the location privacy protection in spatial task allocation. The method was adopted in [24] [25] to protect the publishing of statistical information about location-based datasets guaranteeing that individual location information disclose does not occur. Andres proposed a location perturbation method based on a notion of geo-indistinguishability [26], which is a special application of differential privacy. In [27], the authors used the Markov model to denote the possible locations and protect the exact location with differential privacy.

Recently differential privacy was adopted for task allocation [10], [11] in mobile crowdsensing. In [10], differential privacy mainly protected the aggregated number of workers in a location. However, the trusted third party that computes the aggregated counts of workers has no motivation to participate in practice. Wang et al. [11] avoided involving any third party in the process but the proposed geo-obfuscation function is related to the task locations and need to be updated if the task distribution is transformed, which may discourage the workers who need to download the function before obfuscate their actual locations. Most importantly, both of them employ the same level of privacy protection for all workers, which cannot satisfy the different privacy demands of workers.

2.3 Personalized Privacy

Due to the different data owners have different expectations of privacy protection, personalized privacy has been considered in many privacy-preserving mechanisms, such as personalized privacy for k-anonymity [28]–[30], which allows each user to specify the minimum k they are comfortable with. Recently, Alaggan et al. [31] proposed the concept of heterogeneous differential privacy to preserve the privacy of personal data by considering users' different privacy expecations. Jorgensen et al. proposed the personalized differential privacy to protect the aggregated number of datasets (e.g., count and median) [32], which is not suitable for the location privacy protection.

In this paper, we focus on preserving location privacy of workers in task allocation of mobile crowdsensing, and propose a personalized privacy-preserving task allocation framework, which can allocate tasks effectively while providing personalized location privacy protection.

3 System overview and problem statement

In this section, we first introduce the concept of generalized differential privacy, and then present the proposed framework of the personalized privacy-preserving task allocation in mobile crowdsensing systems based on the concept of generalized differential privacy. Finally, we describe the key problems during the task allocation process: *the winner selection problem* and *the payment determination problem*.

3.1 Generalized Differential Privacy

In [23], the authors proposed that a mechanism K is ϵ-differential private if for any two adjacent databases x, x', and any output Z, the probability distributions $K(x)$, $K(x')$ differ on Z at most by e^ϵ, namely, $K(x)(Z) \leq e^\epsilon K(x')(Z)$. Andres et. al [12] generalized the notion of differential privacy with the Euclidean metric and applied it into scenarios when x, x' are not databases at all, but belong to an arbitrary domain of secrets χ. For instance, it can be applied to deal with geographic locations, which performs as a probabilistic geo-obfuscation process that obfuscates the actual location to another one [26]. For any x, x', if the Euclidean distance $d(x, x') \leq r$, the difference between the distribution $K(x)$ and $K(x')$ should be at most ϵr, where K is the obfuscation mechanism and ϵ denotes the level of privacy at one unit of distance. In this way, the adversary cannot distinguish the actual secrete value of individuals even he/she knows the obfuscation mechanism K. The definition of d_χ-privacy mechanism is as follows.

Definition 1. *(d_χ-privacy [12]) A mechanism K satisfies d_χ-privacy iff for all x, $x' \in \mathcal{X}$:*

$$K(x)(Z) \leq e^{d_{\mathcal{X}}} K(x')(Z)$$

where $K(x)(Z)$ denotes the probability that the reported value belongs to the set $Z \subseteq \mathcal{Z}$. $d_\chi = \epsilon d(x, x')$ and ϵ is the privacy budget, the smaller ϵ, the better privacy protection within $d(x, x')$.

Definition 2. *(Laplace Mechanism [12]) Let $\mathcal{X}$, $\mathcal{Z}$ be two sets where $\mathcal{X} \subset \mathbb{R}$, $\mathcal{Z} = \mathbb{R}$, and let $d_{\mathcal{X}}$ be a metric on $\mathcal{X} \cup \mathcal{Z}$. $D(x)(z) = \frac{\epsilon}{2} \exp\left(-d_{\mathcal{X}}(x, z)\right)$ is a pdf for all $x \in \mathcal{X}$ where $d_{\mathcal{X}} = \epsilon d_{\mathbb{R}}(x, z)$ ($d_{\mathbb{R}}(x, z) = |x - z|$). Then the mechanism K: $\mathcal{X} \to P(\mathcal{Z})$ is called a Laplace mechanism from ($\mathcal{X}$, $d_{\mathcal{X}}$) to $\mathcal{Z}$ and it satisfies $d_{\mathcal{X}}-$privacy.*

Intuitively, when $\mathcal{X}$ and $\mathcal{Z}$ are one-dimensional values, the Laplace Mechanism denotes that the reported value z can be obtained by adding Laplace noise on x where $\lambda = 1/\epsilon$ and $\mu = 0$, and it satisfies $d_{\mathcal{X}}-$privacy where $d_{\mathcal{X}} = \epsilon|x - z|$.

Proposition 1. *If $d_\mathcal{X} \leq d'_\mathcal{X}$, $d_\mathcal{X}$-privacy implies $d'_\mathcal{X}$-privacy [12].*

That is, for a mechanism K which satisfies $d_\mathcal{X}$-privacy, it also satisfies $d'_\mathcal{X}$-privacy if $d_\mathcal{X} \leq d'_\mathcal{X}$.

3.2 System Overview

We consider mobile crowdsensing applications which leverage the power of the crowd to collect massive spatio-temporal information. In general, the typical commercial mobile crowdsensing system consists of three parties: the server, the data requesters and the workers with mobile devices.

It is worth noting that the worker's travel distance is an important metric in task allocation which can be denoted by $d(l_w, l_t)$, where l_w, l_t represent the locations of a worker and a task, respectively, and $d(l_w, l_t)$ is the Euclidean distance between the locations l_w and l_t. Instead of submitting its actual location, a worker can submit the distance between itself and an interested task to the server, so that the server can accurately and efficiently allocate tasks. However, the location privacy can also be leaked by the true distance if the adversary has the prior knowledge about the worker, knowing that the distribution of places $\pi(x)$ which denotes the frequency of worker visiting x. Therefore, in this paper, we propose that each worker uploads the obfuscated distances to its interested tasks to the server, instead of uploading its true location or the true distances to the interested tasks, to protect the worker's location privacy.

Figure 1 shows the proposed personalized privacy-preserving task allocation framework in mobile crowdsensing systems with distance obfuscation. The data requesters create spatial-temporal tasks on the server which then publishes tasks to workers. Workers can obtain the accurate locations of tasks so that they can determine whether to apply for tasks. Instead of applying for the interested tasks using the actual locations or actual distances, the workers can use Laplace mechanism to obfuscate the distance with their personalized privacy budget ϵ. Note that ϵ can be different for different workers which depends on the privacy protection requirement of each worker. The server selects the winner for each task (winner selection process) and gives the winners payments once their uploaded data are certified (payment determination process).

Moreover, we assume that each task can be completed once it is assigned to one worker, unless the worker cannot upload the useful data, which is out of the scope of this paper. A worker can apply for multiple tasks so that it can have more opportunities to be selected as a winner by the server, but it can be assigned at most one task. It is challenging to select winners for tasks and determines the payment for winners as the server is only aware of the obfuscated distances but not the true distances.

3.3 Problem Formulation

We use $W = \{w_1, w_2, \ldots, w_n\}$ to denote the set of workers and $T = \{t_1, t_2, \ldots, t_m\}$ to denote the set of tasks, where n and m are the numbers of workers and tasks, respectively. The data collected by each task $t_j \in T$ has a value v_j for the data requesters and the payment of the task cannot exceed the value. Moreover, each task is associated with a geographical publishing region and only those workers within the region can receive the task information and can apply for it. The publishing region of task t_j is the circular region with the location of t_j as the center and r_j as the radius. Let d_{ij} denote the actual distance of worker w_i to task t_j, while $\tilde{d}_{ij}$ denotes the obfuscated distance obtained by adding the Laplace noise to d_{ij}. Each worker w_i can select some tasks he/she is interested in and then submits the tuples $(t_j, \tilde{d}_{ij})$ and its own privacy budget ϵ_i to the server, where $t_j \in T_i$ is a task in the set of interested tasks of worker w_i. Upon receiving the applications from workers, the server needs to select the winner for each task and determines the payment for each winner w_i.

3.3.1 Winner Selection Problem

The server aims to find the winner for each task, with the obfuscated information, so that the total travel distance for all tasks can be minimized. Let $x(w_i, t_j)$ denote the state of task assignment, where $x(w_i, t_j) = 1$ means that task t_j has been allocated to worker w_i, otherwise $x(w_i, t_j) = 0$. Hence, the winner selection problem is formulated as follows.

$$\begin{aligned} \text{min} \quad & \sum_{t_j \in T} \sum_{w_i \in W} x(w_i, t_j) d_{ij} \\ \text{s.t.} \quad & x(w_i, t_j) = \{0, 1\} \\ & \sum_{t_j \in T} x(w_i, t_j) \leq 1, \quad \forall\, i = 1, 2, \ldots, n \\ & \sum_{w_i \in W} x(w_i, t_j) = 1, \quad \forall\, j = 1, 2, \ldots, m \end{aligned} \tag{1}$$

The objective is to minimize the total travel distance to all tasks. The second constraint indicates that each worker will be assigned to at most one task, and the third constraint indicates that each task will be assigned to one worker.

The formulated problem is an *Integer Linear Programming* problem, which could be easily solved if the actual distances d_{ij} is known. However, in order to protect the location privacy, only the obfuscated distances $\tilde{d}_{ij}$ and personal privacy protection level are uploaded to the server, which makes the winner selection problem difficult to solve.

3.3.2 Payment Determination Problem

The server needs to determine the appropriate payments for the winners by considering the travel costs and privacy protection levels.

If worker w_i is selected as the winner for task t_j. The utility of worker w_i can be denoted by

$$U_{ij} = \begin{cases} P_{ij} - C_{ij}, & \text{if } x(w_i, t_j) = 1 \\ 0, & otherwise \end{cases} \tag{2}$$

where P_{ij} and C_{ij} are the payment and the cost for worker w_i performing task t_j, respectively. The cost C_{ij} is mainly incurred by traveling to the location of t_j (*movement cost* C^m_{ij}) and privacy leakage (*privacy cost* C^p_i). Intuitively, the larger of the travel distance d_{ij}, the larger of the movement cost C^m_{ij}. The larger of the privacy budget ϵ_i, the smaller of the added noise and the larger of the privacy leakage C^p_i. Hence, we have

$$\begin{aligned} C_{ij} &= C^m_{ij} + C^p_i \\ &= \alpha d_{ij} + \beta \epsilon_i \end{aligned} \tag{3}$$

where α and β are coefficients that scale the value of movement cost and privacy cost. The proposed payment determination mechanism should satisfy the following three desirable properties [33]:

- **Individual Rationality**: A worker will have a non-negative utility when performing a task.
- **Profitability**: The value of each task should be at least as large as the payment paid to the winner so that the data requesters can benefit from the data.
- **Truthfulness**: A mechanism is truthful if no worker can improve its utility by cheating its travel distance, no matter what others do.

Note that we assume the workers have no motivation to cheat their privacy budget because the Laplace mechanism will sanitize their travel distance according to their selected privacy budget and directly upload the sanitized distance and privacy budget to server, which makes them cannot get any excessive utility by doing that. It is challenging to design a payment determination mechanism satisfying the above three properties simultaneously without knowing the actual distances of workers.

4 Probabilistic Winner Selection Mechanism

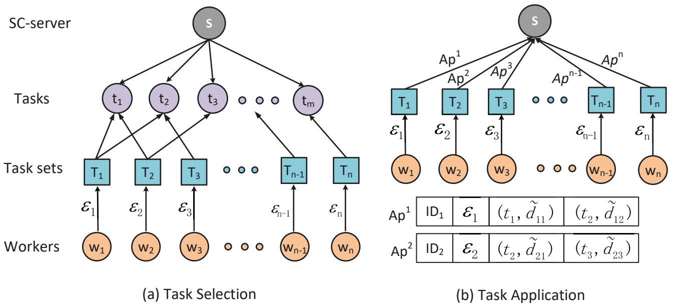
In this section, we focus on the winner selection problem and propose the Probabilistic Winner Selection Mechanism (PWSM) to minimize the total travel distance for task allocation without knowing the true distances to tasks.

As shown in Figure 2(a), each worker can choose a set of tasks from the published tasks. Let Ap^i denote the uploaded application from work w_i that contains the worker ID, privacy budget ϵ_i and (task-distance) tuples. Anonymity or pseudonym techniques can be used for worker ID to protect the Identity information of each user. Note that identity protection is not the focus of this paper, any suitable identity protection techniques can be adopted into our framework. The workers upload the applications as shown in Figure 2(b). Upon receiving Ap^i, the server aims to find the winner for each task so that the total travel distance can be minimized.

However, without knowing the true distances, it is difficult to find the optimal task allocation to minimize the total travel distance. To solve this problem, we propose a suboptimal solution that finds the winner of a task as the worker who has the largest probability of being closest to the task. In particular, some workers may be selected as the winners of multiple tasks, so we also propose a conflict elimination algorithm (CEA) to solve the conflicts. In the following, we introduce the proposed mechanisms starting from the single task scenario, and then moving to the multiple tasks scenario.

4.1 Single Task Scenario

We start from the single task scenario where the server publishes only one task t_k and there are multiple workers applying for it. The objective is to find the closest worker to the task, given the obfuscated distances and personal privacy budgets, which is difficult as the true distances are unknown. To solve this problem, we transform the problem to finding the worker with the largest probability of being closest to the task.



(a) Task Selection

(b) Task Application

Fig. 2. The selection and application process of workers in our proposed task allocation model.

Let us start from the special case that only two workers w_i and w_j applying for the task, and then extend it to a more general case. Worker w_i uploads ϵ_i and $\tilde{d}_{ik}$, and worker w_j uploads ϵ_j and $\tilde{d}_{jk}$. The server aims to obtain the probability of d_{ik} smaller than d_{jk}, denoted by $P(d_{ik} \leq d_{jk})$. The worker w_i gets the sanitized distance $\tilde{d}_{ik}$ by adding the Laplace noise on d_{ik}, hence we have

$$d_{ik} = \tilde{d}_{ik} - \eta_i, \qquad \eta_i \sim \text{Laplace}(0, 1/\epsilon_i) \tag{4}$$

Similarly we have

$$d_{jk} = \tilde{d}_{jk} - \eta_j, \qquad \eta_j \sim \text{Laplace}(0, 1/\epsilon_j) \tag{5}$$

where η_i and η_j are variables that follow the Laplace distribution. Note that the smaller of ϵ_i, the larger of the added noise, and the stronger the privacy protection level. Then we have

$$\begin{aligned} P(d_{ik} \leq d_{jk}) &= P(\tilde{d}_{ik} - \eta_i \leq \tilde{d}_{jk} - \eta_j) \\ &= P(\tilde{d}_{ik} - \tilde{d}_{jk} \leq \eta_i - \eta_j) \end{aligned} \tag{6}$$

where $\tilde{d}_{ik} - \tilde{d}_{jk}$ is known by the server. The above equation can be seen as a probability problem about two-dimensional continuous variables (η_i, η_j) in the plane set D, denoted by $P((\eta_i - \eta_j) \in D)$. The plane D can be denoted by

$$D = \{(\eta_i, \eta_j) : \eta_i - \eta_j \geq \tilde{d_{ik}} - \tilde{d_{jk}}\} \tag{7}$$

The double integral operation can be used for solving this problem, we have

$$\begin{aligned} P(\tilde{d}_{ik} - \tilde{d}_{jk} \leq \eta_i - \eta_j) &= P((\eta_i - \eta_j) \in D) \\ &= \iint_D f(\eta_i, \eta_j) d\eta_i d\eta_j \end{aligned} \tag{8}$$

where $f(\eta_i, \eta_j)$ denotes the *joint probability density function* of (η_i, η_j). Since the variables η_i, η_j are independent from each other, we have

$$\begin{aligned} \iint_D f(\eta_i, \eta_j) d\eta_i d\eta_j &= \iint_D f(\eta_i) f(\eta_j) d\eta_i d\eta_j \\ &= \int_{-\infty}^{\infty} \Big(\int_{-\infty}^{\eta_i - (\tilde{d}_{ik} - \tilde{d}_{jk})} f(\eta_i) f(\eta_j) d\eta_j \Big) d\eta_i \end{aligned} \tag{9}$$

Note that

$$f(\eta_i) f(\eta_j) = \frac{\epsilon_i \epsilon_j}{4} e^{-(\epsilon_i |\eta_i| + \epsilon_j |\eta_j|)} \tag{10}$$

Equation (9) can be seen as a function, where the inputs of the function are $(\tilde{d}_{ik}, \tilde{d}_{jk}, \epsilon_i, \epsilon_j)$, and the output is the probability of d_{ik} smaller than d_{jk}. We call this function as the *Probability Compare Function* (PCF). Combining Equations (6) and (9), we have

$$\begin{aligned} P(d_{ik} \leq d_{jk}) &= PCF(\tilde{d_{ik}}, \tilde{d_{jk}}, \epsilon_i, \epsilon_j) \\ &= \iint_D f(\eta_i, \eta_j) d\eta_i d\eta_j \end{aligned} \tag{11}$$

If $P(d_{ik} \leq d_{jk}) > 1/2$, we can say that worker w_i will be closer than worker w_j with a larger probability. Based on this principle, we can compare any two workers who apply the task and finally find the worker who has the largest probability of being closest to the task.

4.2 Multiple Tasks Scenario

Next we focus on the multiple tasks scenario. As shown in Figure 2, we assume there are n workers applying for m tasks. For each task, there are at most n workers applying for it. To solve the winner selection problem, for each task, we utilize the proposed PCF to determine which worker is closer and sort them in descending order in terms of probability. We use a matrix $A_{m\times n}$ to denote the sorted indices of workers for all tasks, which is

$$A_{m\times n} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \tag{12}$$

where a_{ij} is the index of a worker, $a_{ij} \in \{1, 2, \ldots, n\}$. For example, $a_{ij} = k$ means that worker w_k who applies task t_i is ranked at the jth position. If there are only ρ ($\rho \leq n$) workers applying for task t_i, a_{ij} will be ∞ for any $j > \rho$.

Let A_i denote the ith row of matrix A, which is the sorted indices of workers who apply for task t_i. Let $S = [s_1, s_2, \ldots, s_m]$ denote the set of winners for all tasks, where s_i denotes the winner for task t_i. In general, the expected winner should be $A_i[1]$ for task t_i, that is, $s_i = A_i[1] = a_{i1}$. If each expected winner is allocated with at most one task, the winner selection process is completed.

However, there may exist some workers who are selected as the winners of multiple tasks, while each worker can be allocated with at most one task. We call this as the *winner conflict problem*. Given the vector S, the winner conflict happens if there exist i and j satisfying $S[i] = S[j]$.

4.2.1 Conflict Elimination Algorithm (CEA)

We propose the CEA to solve the winner conflict problem. The key idea is that, for any conflict, say w_c is selected as the winner of φ tasks, we allocate only one task to w_c and find another candidate other than w_c for each of the rest $\varphi - 1$ conflicted tasks, with the objective of minimizing the total travel distance. We repeat this process until there is no conflict in the final winner vector S.

In the following, we use an example shown in Figure 3 to better explain the conflict elimination process when worker w_c is selected as the winners by φ tasks. The best candidate beside w_c is the one who has the second largest probability of being closest to the each task. That is, the candidate for task t_i is a_{i2} if $a_{i1} = c$. Note that this example only shows how to eliminate the conflicts of w_c, and new conflicts may happen for the new winners, and the proposed CEA will iteratively solve the conflict until no conflict exists. For the example in Figure 3, we present all possible conflicts elimination situations for w_c as follows.

$$\begin{cases} \mathcal{C}_1 : D_i = D(a_{i1}) + D(a_{j2}) + \ldots + D(a_{k2}) \\ \mathcal{C}_2 : D_j = D(a_{i2}) + D(a_{j1}) + \ldots + D(a_{k2}) \\ \qquad\qquad \vdots \\ \mathcal{C}_\varphi : D_k = D(a_{i2}) + D(a_{j2}) + \ldots + D(a_{k1}) \end{cases} \tag{13}$$

where $D(a_{ij})$ denotes the actual travel distance for the worker with index a_{ij} to task t_i. D_i denotes the travel distance for the φ conflicted tasks when t_i is assigned to w_c and the other tasks are assigned to the corresponding candidates who rank behind the winners. The objective is to find the minimum travel distance from all situations in Equation (13). Hence, we need to compare the travel distance for any two situations.

$$D_i - D_j = D(a_{i1}) + D(a_{j2}) - D(a_{i2}) - D(a_{j1}) \tag{14}$$

Since $a_{i1} = a_{j1} = c$, $D(a_{i1})$ and $D(a_{j1})$ can be replaced by d_{ci} and d_{cj}, respectively. We assume that a_{i2} denotes the index of w_a, a_{j2} denotes the index of w_b. Then $D(a_{i2})$ and $D(a_{j2})$ can be replaced by d_{ai} and d_{bj}, respectively. Hence the above Equation can be represented by

$$D_i - D_j = d_{ci} + d_{bj} - d_{ai} - d_{cj} \tag{15}$$

However, the server cannot obtain the actual travel distance of workers and only knows the sanitized distance $\tilde{d}_{ci}$, $\tilde{d}_{bj}$, $\tilde{d}_{ai}$ and $\tilde{d}_{cj}$. Therefore, the objective to compare D_i and D_j will be converted to find the probability of $D_i - D_j \leq 0$.

$$\begin{aligned} P(D_i - D_j \leq 0) &= P(d_{ci} + d_{bj} - d_{ai} - d_{cj} \leq 0) \\ &= P(d_{ci} - d_{cj} \leq d_{a_i} - d_{bj}) \end{aligned} \tag{16}$$

Based on Equation (6), we have

$$\begin{aligned} & P(d_{ci} - d_{cj} \leq d_{ai} - d_{bj}) = P(d_{ci} + d_{bj} - d_{ai} - d_{cj} \leq 0) \\ & = P(\tilde{d}_{ci} - \eta_{c_1} - \tilde{d}_{cj} + \eta_{c_2} \leq \tilde{d}_{ai} - \eta_a - \tilde{d}_{bj} + \eta_b) \\ & \eta_{c_1} \sim \text{Laplace}(0, 1/\epsilon_c), \quad \eta_{c_2} \sim \text{Laplace}(0, 1/\epsilon_c) \\ & \eta_a \sim \text{Laplace}(0, 1/\epsilon_a), \quad \eta_b \sim \text{Laplace}(0, 1/\epsilon_b) \end{aligned} \tag{17}$$

where there exist four variables η_{c_1}, η_{c_2}, η_a and η_b, so that we cannot get the result until more constraints about those variables are known. To solve this problem, we roughly assume that the difference between the travel distances for different tasks is relatively small if those tasks are applied for by the same worker. That is, d_{ci} can be thought close to d_{cj}. Hence we have

$$\begin{aligned} P(D_i \leq D_j) &= P(d_{ai} - d_{bj} \geq 0) \\ &= P(d_{ai} \geq d_{bj}) \end{aligned} \tag{18}$$

$P(D_i \leq D_j)$ can be calculated by the proposed PCF. If $P(d_{ai} \geq d_{bj}) \geq 1/2$, D_i has a larger probability of being smaller than D_j, so task t_i should be assigned to w_c and task t_j will be assigned to the candidate.

We can see that the problem to compare D_i and D_j has been converted to compare d_{ai} and d_{bj}, which are travel

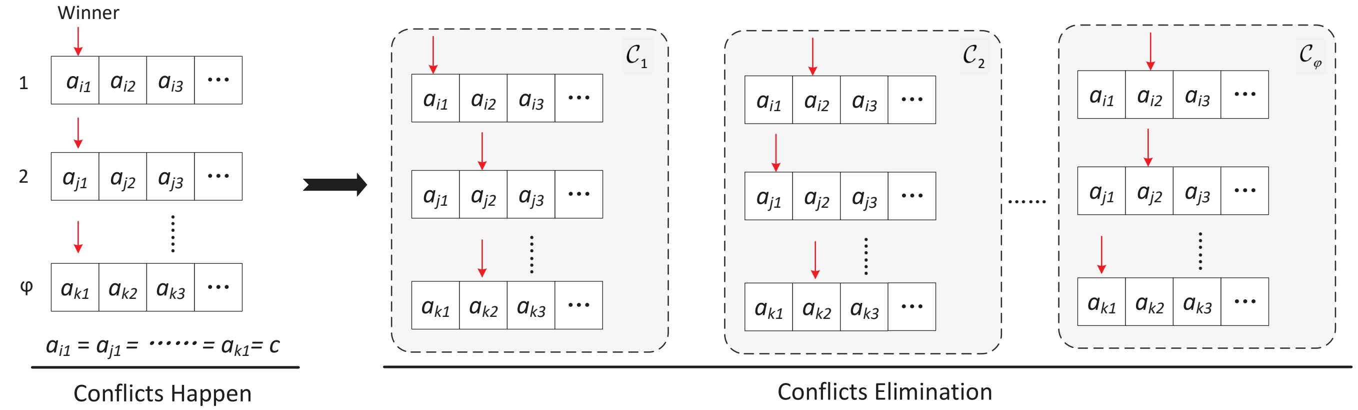

Winner
1
2
φ
a_{i1} a_{i2} a_{i3} ···
a_{j1} a_{j2} a_{j3} ···
a_{k1} a_{k2} a_{k3} ···
$a_{i1} = a_{j1} = \cdots = a_{k1} = c$
Conflicts Happen
$\mathcal{C}_1$
$\mathcal{C}_2$
$\mathcal{C}_\varphi$
Conflicts Elimination

Fig. 3. Illustration of the conflict elimination process when worker w_c is selected as the winners of φ tasks. a_{i1} denotes the expected winner of task t_i after sorting, and a_{i2} is the candidate with the second largest probability of being closest to t_i.

distances of candidates for task t_i and task t_j, respectively. The comparison for all the conflicts can be compared similarly. Note that conflicts would not happen commonly in practice because tasks are usually distributed sparsely. Moreover, the number of tasks each worker can apply for are limited, which further reduces the occurrence of winner conflict.

4.2.2 PWSM

The key idea of PWSM is firstly using the PCF to find a winner for each task from workers who apply for it, and then using the CEA to eliminate the winner conflicts. The procedures are descried as follow:

1) Use the PCF to construct the matrix $A_{m\times n}$ that denotes the sorted probabilities of workers for m tasks.
2) Pick up the first element in each row as the winner for each task, and put them into S.
3) If S contains the same element, winner conflict happens and repeatedly apply the CEA to eliminate conflicts until there is no conflict in S.

Theorem 2. *The complexity of the proposed PWSM is $O(\kappa^2\frac{n^2}{m}+\kappa^2 n)$, where m and n are the number of tasks and workers, respectively, and κ is the largest number of tasks a worker can apply for.*

Proof. The PWSM mainly contains the sorting process and the conflict elimination process. The average number of applications each task has is $\kappa n/m$, so that sorting the workers with the typical bubble sorting algorithm takes $O((\kappa n/m)^2)$ for each task. For all m tasks, the complexity of obtaining $A_{m\times n}$ will be $O(\kappa^2\frac{n^2}{m})$. As for the conflict elimination process, the worst-case to eliminate conflicts is that it takes $O(\kappa(\kappa n/m)$ for each task to eliminate conflicts. Since there are m tasks, the complexity of CEA will be $O(\kappa^2 n)$. Hence the complexity of PWSM is $O(\kappa^2\frac{n^2}{m}+\kappa^2 n)$. □

5 Vickrey Payment Determination Mechanism

In this section, we propose the vickrey payment determination mechanism (VPDM) to determine the appropriate payment by considering its travel distance and privacy level, which should satisfy the truthfulness, profitability and probabilistic individual rationality.

In payment determination, the payment paid to the winner cannot exceed the value of the task itself. That is the payment P_{ij} of winner w_i for task t_j cannot exceed the task value v_j. The server can use the parameters α and β to control the payment and use them to assess the value of moving distance and privacy. For example, if the winner w_i applies for task t_j with the privacy budget ϵ_i and the actual distance is d_{ij}. The cost C_{ij} of w_i for performing task t_j is

$$C_{ij} = \alpha d_{ij} + \beta\epsilon_i \tag{19}$$

where the larger of ϵ_i, the smaller of the added noise and the larger of the privacy leakage. Since the cost is incurred by two aspects, we determine the payment from compensating the movement cost and privacy cost, respectively. We have

$$P_{ij} = P_{ij}^m + P_{ij}^l \tag{20}$$

where P_{ij} is the payment for winner w_i performing task t_j. P_{ij}^m and P_{ij}^l denote the payment for the movement cost and privacy leakage, respectively.

We use r_j to denote the radius of the task publishing region, hence the true distances of workers who apply for task t_j cannot exceed r_j. Let ϵ_{max} denote the largest privacy budget that workers can use for protecting their locations. Then we have

$$\begin{cases} \alpha_j r_j + \beta_j\epsilon_{max} = v_j \\ \alpha_j = \kappa\beta_j \end{cases} \tag{21}$$

where κ is the parameter to scale the difference of α_j and β_j. In order to guarantee that the cost of all the tasks cannot exceed their values, we have

$$\alpha = \min(\alpha_i), \quad \beta = \min(\beta_j), \quad i,j \in [1,m] \tag{22}$$

As for the payment P_{ij}^l, the server can straightly pay the winner according to the submitted privacy budget ϵ_i and parameter β. It's simple to see that the $P_{ij}^l = \beta\epsilon_i$ will satisfy the individual rationality and truthfulness, because the utility of workers is always 0 and can be non-negative. Moreover, the workers have no motivation to cheat their privacy budget because the uploaded sanitized distances are generated by the selected privacy budget.

While for the determination of the payment P_{ij}^m, it is challenging to guarantee the individual rationality and truthfulness because d_{ij} is only known to the worker itself. Hence the server aims to find $\hat{d}_{ij}$ ($\hat{d}_{ij} \geq d_{ij}$) and then calculate $P_{ij}^m = \alpha\hat{d}_{ij}$, so that the utility of a worker would not be negative. From Equation (4), we have

$$P(\hat{d}_{ij} - d_{ij} \geq 0) = P(\hat{d}_{ij} - \tilde{d}_{ij} + \eta_i \geq 0) \tag{23}$$

where $\eta_i \sim$ Laplace$(0, 1/\epsilon_i)$. Hence the server can use the above equation to find $\hat{d}_{ij}$ that guarantees the probability of $\hat{d}_{ij} - d_{ij} \geq 0$ at least be $\mathcal{P}$.

Definition ($\mathcal{P}$-Individual Rationality) A Mechanism M satisfies $\mathcal{P}$-Individual Rationality if each worker has a non-negative utility with a probability of at least $\mathcal{P}$.

Suppose we find a $\hat{d}_{ij}$, $P_{ij} = \alpha\hat{d}_{ij} + \beta\epsilon_i$ satisfies the $\mathcal{P}$-Individual Rationality, however, it may not satisfy the truthfulness. Worker w_i who applies for task t_j can calculate $\tilde{d}_{ij}$ using a fake distance $\bar{d}_{ij}$ instead of true distance d_{ij}, where $\min(d_{-ij}) > \bar{d}_{ij} > d_{ij}$. d_{-ij} denotes the distance of other workers who apply task t_j except worker w_i. We can see that worker w_i can still be the winner and acquire more payment by cheating his/her actual distance. Inspired by the *Vickrey Auction* (a truthful auction mechanism) where the highest bidder wins but the price paid is the second-highest bid, we use the candidate's distance to determine the payment for the winner. Suppose the candidate for task t_j is w_c, then we can use d_{cj} to replace d_{ij} in Equation (23), and we have

$$P(\hat{d}_{ij} - d_{cj} \geq 0) \geq \mathcal{P} \tag{24}$$

where d_{cj} is the actual distance of candidate w_c for task t_j.

According to Equation (4) and the symmetry of Laplace distribution, we have

$$d_{cj} = \tilde{d}_{cj} + \eta_c, \qquad \eta_c \sim \text{Laplace}(0, 1/\epsilon_c) \tag{25}$$

Then we have

$$d_{cj} \sim \text{Laplace}(\tilde{d}_{cj}, 1/\epsilon_c) \tag{26}$$

The Equation (24) can be converted to

$$\begin{aligned} P(\hat{d}_{ij} - d_{cj} \geq 0) \geq \mathcal{P} &\Leftrightarrow P(d_{cj} \leq \hat{d}_{ij}) \geq \mathcal{P} \\ &\Leftrightarrow F(\hat{d}_{ij}) \geq \mathcal{P} \end{aligned} \tag{27}$$

where $F(x)$ is the distribution function of d_{cj} and can be represented by

$$F(x) = \int_{-\infty}^{x} \frac{\epsilon_c}{2} e^{-(\epsilon_c|x - \tilde{d}_{cj}|)} \tag{28}$$

Note that the server needs to give more payment to the winner for a larger $\hat{d}_{ij}$. In order to save the cost of the server, $\hat{d}_{ij}$ should be the threshold that satisfies $F(\hat{d}_{ij}) = \mathcal{P}$, which can be calculated based on Equation (27) and Equation (28), given that ϵ_c and $\tilde{d}_{cj}$ are known by the server. Moreover, $\hat{d}_{ij}$ should not be larger than r_j. If no candidate can be found, we also let $\hat{d}_{ij} = r_j$. Hence the payment P_{ij} of worker w_i performing task t_j is

$$P_{ij} = \alpha\hat{d}_{ij} + \beta\epsilon_i \tag{29}$$

where α and β are calculated from Equation (22), and $\hat{d}_{ij}$ is calculated from $F(\hat{d}_{ij}) = \mathcal{P}$.

Theorem 3. *The proposed VPDM is $\mathcal{P}P(d_{cj} \geq d_{ij})$-Individual Rationality.*

Proof. Based on Equation (3) and Equation (29), the utility of worker w_i for task t_j is

$$\begin{aligned} U_{ij} &= (\alpha\hat{d}_{ij} + \beta\epsilon_i) - (\alpha d_{ij} + \beta\epsilon_i) \\ &= \alpha(\hat{d}_{ij} - d_{ij}) \end{aligned} \tag{30}$$

Therefore, the probability of the winner receiving a non-negative utility can be converted to the probability of $\hat{d}_{ij}$ larger than d_{ij}, which is

$$\begin{aligned} P(\hat{d}_{ij} - d_{ij} \geq 0) &\geq P(\hat{d}_{ij} - d_{cj} \geq 0)P(d_{cj} \geq d_{ij}) \\ &\geq \mathcal{P}P(d_{cj} \geq d_{ij}) \end{aligned} \tag{31}$$

That is, each winner has a non-negative utility with a probability of at least $\mathcal{P}P(d_{cj} \geq d_{ij})$, where $P(d_{cj} \geq d_{ij})$ can be calculated by the PCF. □

Theorem 4. *The proposed VPDM is truthful.*

Proof. The VPDM can be truthful if and only if the winner cannot improve its utility by unilaterally cheating its true distance (e.g., adds Laplace noise to a fake distance $\bar{d}_{ij}$ instead of the true distance d_{ij}). We discuss the problem from the following situations:

- ($\bar{\boldsymbol{d}}_{\boldsymbol{ij}} < \boldsymbol{d}_{\boldsymbol{ij}}$): If $d_{ij} < \min(d_{-ij})$, w_i is still the winner and the payment does not change. If $d_{ij} > \min(d_{-ij}) > \bar{d}_{ij}$, w_i becomes the winner by cheating but the payment is decided by $\min(d_{-ij})$, so the liar will obtain a negative utility.
- ($\boldsymbol{\min(d_{-ij})} > \bar{\boldsymbol{d}}_{\boldsymbol{ij}} > \boldsymbol{d}_{\boldsymbol{ij}}$): Worker w_i is still the winner but his/her utility does not change.
- ($\bar{\boldsymbol{d}}_{\boldsymbol{ij}} > \boldsymbol{\min(d_{-ij})}$): If $d_{ij} < \min(d_{-ij})$, worker w_i was the winner but now is not the winner after cheating, so its utility becomes 0. If $d_{ij} > \min(d_{-ij})$, w_i is always not selected as the winner.

Based on the above analysis, we can see that workers cannot improve their utility by unilaterally cheating its true distance, which means that workers will be truthful in the process of task allocation. □

Theorem 5. *The proposed VPDM is profitable.*

Proof. We have $\hat{d}_{ij} \leq r_j$ and $\epsilon_i \leq \epsilon_{max}$. According to Equation (21), we have

$$\alpha_j\hat{d}_{ij} + \beta_j\epsilon_i \leq v_j \tag{32}$$

Since $\alpha \leq \alpha_j$ and $\beta \leq \beta_j$, we have

$$p_{ij} = \alpha\hat{d}_{ij} + \beta\epsilon_i \leq v_j \tag{33}$$

Therefore, the proposed payment determination mechanism is profitable. □

6 Privacy Analysis

Theorem 6. *Let $X_i \in R^k$ denote the set of actual distances of worker w_i to its interested tasks T_i where $|T_i| = k$, and M be a mechanism where $M(X_i) = X_i + Lap(1/\epsilon_i)$. Then, M satisfies $\epsilon_i \sum_{t_j \in T_i} r_j$-privacy where r_j is the publishing region of t_j.*

Proof. For any $X_i, X'_i \in R^k$ where $x_{ij} \in X_i$, $x'_{ij} \in X'_i$ denote the actual distance of worker w_i to task t_j and $d(x_{ij}, x'_{ij}) \leq r_j$. $Z_i \in R^k$ denotes the set of reported distances of worker w_i to tasks in T_i. $Z_i = X_i + (\eta_1, \eta_2, \ldots, \eta_k)$ where η_j are i.i.d random variables drawn from $Lap(1/\epsilon_i)$. Hence we have,

$$\begin{aligned}\frac{P(X_i = Z_i)}{P(X'_i = Z_i)} &= \prod_{t_j \in T_i} \left(\frac{\exp(-\epsilon_i|z_{ij} - x_{ij}|)}{\exp(-\epsilon_i|z_{ij} - x'_{ij}|)}\right) \\ &= \prod_{t_j \in T_i} \exp(\epsilon_i(|z_{ij} - x_{ij}| - |z_{ij} - x'_{ij}|)) \\ &\leq \prod_{t_j \in T_i} \exp(\epsilon_i|x_{ij} - x'_{ij}|) \\ &= \exp(\epsilon_i\|X_i - X'_i\|_1)\end{aligned} \tag{34}$$

Note that each task t_j is associated with a geographical publishing region with the radius of r_j. Hence the $\|X_i - X'_i\|_1 \leq \sum_{j=1}^{k} r_j$. Hence M satisfies $\epsilon_i \sum_{t_j \in T_i} r_j -$ privacy. □

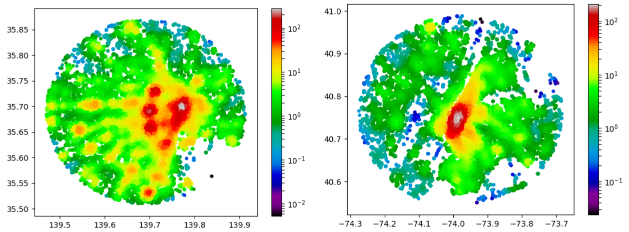
From Theorem 6, we can see that different workers can have different privacy levels and the privacy level of each worker is related to its selected privacy budget and applied tasks. The smaller of the privacy budget ϵ_i, the better of the privacy protection. The smaller of the number of applied tasks and r_j, the higher of the privacy protection. This is because more information about the location of worker w_i will be exposed with the increase of the number of applied tasks, which means higher probability of privacy leakage. Moreover, $d(x_{ij}, x'_{ij})$ will increase with the increase of r_j, which means higher distinguishability or lower privacy protection. Note that the location of worker w_i would not be leaked even with a very large r_j. For instance, when $r_j = 1000$ km, $d_{\mathcal{X}}$ becomes large so that the sever can infer whether the worker w_i is located in Paris or London with a high probability, but it still cannot infer the actual location of w_i in the city.

7 Experimental Evaluation

In this section, we evaluate the performance of the proposed PWSM and VPDM on a real-world check-in dataset. We implement the mechanisms in Python and compare its performance with two benchmark mechanisms.

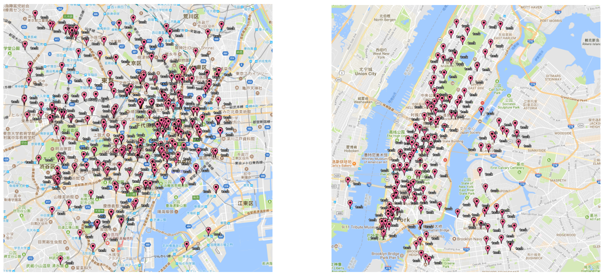
7.1 Dataset

We use the dataset collected from Foursquare in [34], which includes long-term (about 10 months) check-in data in New York and Tokyo from 12 April 2012 to 16 February 2013. The data of New York contains 227428 check-ins and Tokyo contains 573708 check-ins. The density of check-ins in 10 months is shown in Figure 4. We can see that most of the check-in events happened in the central urban areas, hence we distribute the tasks in these areas and users in these areas can be seen as the workers. The tasking area of Tokyo is within longitude (139.68, 139.80) and latitude (35.62, 35.74), and the tasking area of New York is within longitude (-74.04, -73.93) and latitude (40.69, 40.80).



(a) Tokyo (b) New York

Fig. 4. The density of check-ins in two cities.



(a) Tokyo (b) New York

Fig. 5. The distribution of tasks on the real map in two cities.

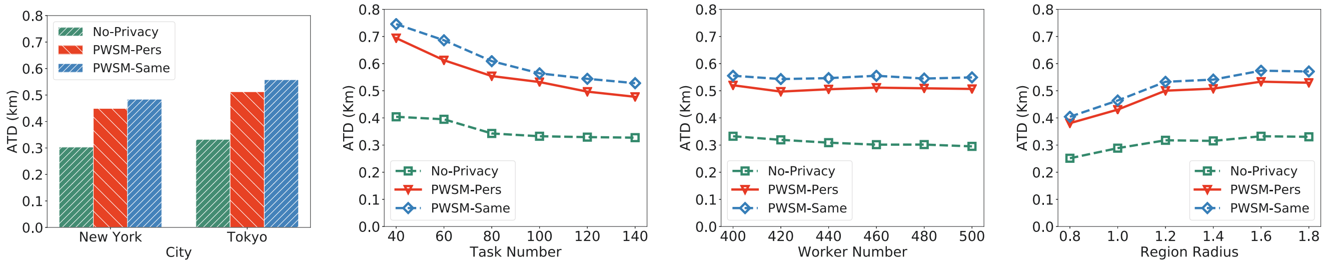
7.2 Setup and Metrics

In order to evaluate the winner selection accuracy of the proposed mechanisms, we conduct experiments on datasets from two representative cities Tokyo and New York. The data collection scenario can be taking pictures at the subway stations. We set the locations of tasks based on the subway stations and the locations of workers based on the office locations of users in datasets. There are 325 subway stations in Tokyo and 142 in New York, and there are 503 offices in Tokyo and 492 in New York. Note that the number of offices is the result after removing the situation that each user ID has multiple offices with different locations.

In the experiments, we randomly select a certain number of subway stations as the locations of tasks and office locations as the locations of workers. First, we evaluate the performance of the proposed mechanisms based on the different task distribution shown in Figure 5 when there are 100 tasks and 400 workers. Second, we mainly use the Tokyo dataset to investigate how our proposed mechanisms perform when the different key parameters (e.g., task number) vary. The number of tasks ranges from 40 to 140 and the number of workers ranges from 400 to 500. The publishing region r_i ranges from 0.8km to 1.8km and the default value is 1.5km. The privacy budget of workers ranges from 1 to 5. Each worker will apply for at most 3 tasks that are closest to him/her.

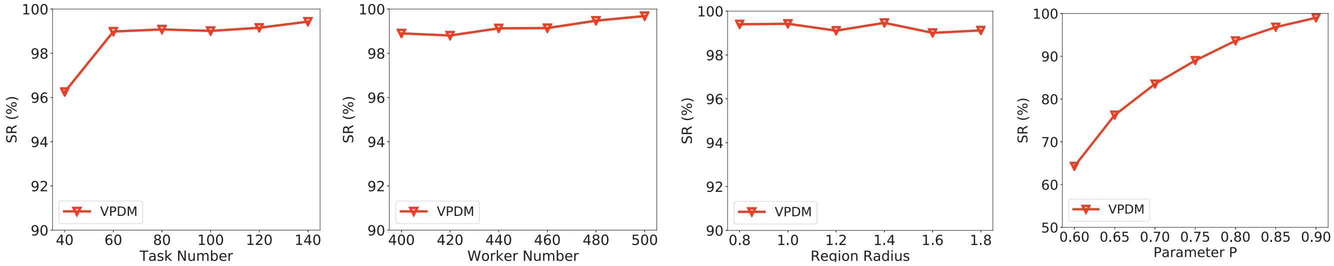
We use two metrics to evaluate the effectiveness of the proposed mechanisms: the average travel distance (ATD) for PWSM and the satisfactory rate (SR) for VPDM.

- ATD: the real total travel distance of the winners divided by the number of successfully allocated tasks.



(a) distribution (b) task number (c) worker number (d) publishing region

Fig. 6. Comparison of the winners selection mechanisms on the average travel distance.



(a) task number (b) worker number (c) publishing region (d) parameter $\mathcal{P}$

Fig. 7. The performance of payment determination mechanism on the satisfactory rate.

- SR: the ratio of the number of winners who receive the non-negative payment to the number of total winners.

We compare PWSM with two benchmarks.

- No-Privacy: The optimal winner selection mechanism when workers' real locations are reported to the server without any privacy protection.
- Same-Privacy (PWSM-Same): The winner selection mechanism based on the PWSM when the privacy budgets of all workers are the same. We set the privacy budget ϵ as 1 for all workers.

7.3 Evaluation on PWSM

Figures 6 shows the performance comparison of three winner selection mechanisms on the average travel distance (ATD) under different parameters. "PWSM-Pers" means that each worker randomly selects its personal privacy level ϵ_i from $[1, 5]$, while "PWSM-Same" means that all workers adopt the same privacy level of $\epsilon = 1$.

Figure 6(a) shows the ATD of three winner selection mechanisms against the city when there are 400 workers and 100 tasks. We can see that workers take a larger ATD to finish tasks in Tokyo than New York. This is because the task distribution in Tokyo is sparser than that in New York. Figure 6(b) shows the ATD of three winner selection mechanisms against the number of tasks when there are 400 workers. We can see that the ATD decreases with the increase of task number, which is because the average travel distance to tasks is smaller with higher density of tasks. Figure 6(c) shows the ATD of three winner selection mechanisms against the number of workers when there are 100 tasks. We can see that the ATD does not change significantly with the increase of worker number. Figure 6(d) shows the ATD of three winner selection mechanisms against the radius of publishing region when there are 400 workers and 100 tasks. We can see that the ATD increases for larger size of the task publishing regions. This is because that, when the size of the publishing region increases, some far away tasks that were not allocated to any worker are now allocated to workers. Since the distances between the workers and far away tasks are large, the ATD of all workers increases when the size of publishing region increases.

From Figure 6, we can see that the PWSM always has a larger ATD than No-privacy, which is because the latter can optimally allocate tasks to workers with the true locations of workers. However, PWSM provides strongly personalized location privacy protection for workers. We can also observe that the ATD of PWSM with different protection levels is smaller than that of PWSM with the same largest protection level, which is because the true distances are obfuscated with larger noise when the strongest protection level is adopted by all workers.

7.4 Experiments for VPDM

Figures 7(a-c) show the performance of the proposed VPDM on the satisfactory rate (SR) when $\mathcal{P}$ is 0.9. We can see that the VPDM always has a SR higher than 96%, which is larger than $\mathcal{P} = 0.9$. Figure 7(a) shows the SR against the number of tasks when there are 400 workers. We can see that the SR increases from 96% to 99% when task number increases from 40 to 60, and thereafter SR does not increase significantly. This is because the number of total winners is small when there are 40 tasks so that one or two winner who receive the negative payment will decrease the SR sharply. Figure 7(b) shows the SR against the number of workers when there are 100 tasks. We can see that the SR increases slightly with the increase of the number of workers. This is because the effect of winners who receive the negative payment will decrease with the increase of the number of total winners. Figure 7(c) shows the SR against the radius of the publishing region when there are 100 tasks and 400 workers. We can see that the SR does not change with the increase of publishing region size.

Figure 7(d) shows the SR against the parameter $\mathcal{P}$ when there are 100 tasks and 400 workers. We can see that the SR increases with the increase of parameter $\mathcal{P}$. This is because workers has a larger probability to get a non-negative payment with the increase of parameter $\mathcal{P}$. We can also observe that SR is always larger than $\mathcal{P}$, which validates the correctness of the proposed payment determination mechanism.

8 Conclusions

In this paper, we proposed a personalized privacy-preserving task allocation framework for mobile crowdsensing systems. Each worker uploads the obfuscated distances and personal privacy budget to the server instead of uploading its true location or true distances to tasks. In particular, we proposed the PWSM that allocates tasks to workers with only the obfuscated information while minimizing the total travel distance. We also proposed the VPDM to determine the appropriate payment to each winner by considering its movement cost and privacy leakage. We proved that the proposed framework provides personalized privacy protection, and satisfies the truthfulness, profitability and probabilistic individual rationality. Moreover, we proved that each worker can have $\epsilon_i \sum_{t_j \in T_i} r_j$-privacy in our framework, which is related to its personal privacy budget and interested tasks. Extensive experiments on the read-world datasets demonstrate the effectiveness of the proposed mechanisms.

Acknowledgments

This work was supported in part by National Natural Science of China (Grants No. 61502352, 61772551, U1636219 and 61373167), the National Basic Research Program of China (Grant 2014CB340600), National Science Foundation (Grants No. 1444059, 1717315), Shandong Provincial Key Program of Research and Development (2018GGX101035), Natural Science Foundation of Hubei Province (Grants No. 2017CFB503, 2017CFA007 and 2017CFA047), Fundamental Research Funds for the Central Universities (Grant No. 2042018gf0043, 413000035 and 18CX07003A), and State Key Lab. for Novel Software Technology, Nanjing University (Grant No. KFKT2018B09). Qian Wang is the corresponding author.

References

[1] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, “Common sense: participatory urban sensing using a network of handheld air quality monitors,” in *Proc. of ACM SenSys*, 2009, pp. 349–350.

[2] T. Aitamurto, “The impact of crowdfunding on journalism: case study of spot. us, a platform for community-funded reporting,” *Journalism practice*, vol. 5, no. 4, pp. 429–445, 2011.

[3] S. B. Liu and L. Palen, “The new cartographers: Crisis map mashups and the emergence of neogeographic practice,” *Cartography and Geographic Information Science*, vol. 37, no. 1, pp. 69–90, 2010.

[4] D. C. Brabham, “Crowdsourcing the public participation process for planning projects,” *Planning Theory*, vol. 8, no. 3, pp. 242–262, 2009.

[5] “Waze,” https://www.waze.com/.

[6] B. Liu, W. Zhou, T. Zhu, H. Zhou, and X. Lin, “Invisible hand: a privacy preserving mobile crowd sensing framework based on economic models,” *IEEE Transactions on Vehicular Technology*, vol. 66, no. 5, pp. 4410–4423, 2017.

[7] C. Cornelius, A. Kapadia, D. Kotz, D. Peebles, M. Shin, and N. Triandopoulos, “Anonysense: privacy-aware people-centric sensing,” in *Proc. of ACM Mobisys*, 2008, pp. 211–224.

[8] L. Pournajaf, L. Xiong, V. Sunderam, and S. Goryczka, “Spatial task assignment for crowd sensing with cloaked locations,” in *Proc. of IEEE MDM*, vol. 1, 2014, pp. 73–82.

[9] I. J. Vergara-Laurens, D. Mendez, and M. A. Labrador, “Privacy, quality of information, and energy consumption in participatory sensing systems,” in *Proc. of IEEE PerCom*, 2014, pp. 199–207.

[10] H. To, G. Ghinita, and C. Shahabi, “A framework for protecting worker location privacy in spatial crowdsourcing,” *Proc. of the VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.

[11] L. Wang, D. Yang, X. Han, T. Wang, D. Zhang, and X. Ma, “Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation,” in *Proc. of WWW*, 2017, pp. 627–636.

[12] K. Chatzikokolakis, M. E. Andrés, N. E. Bordenabe, and C. Palamidessi, “Broadening the scope of differential privacy using metrics,” in *Proc. of PETS*, 2013, pp. 82–102.

[13] B. Guo, Y. Liu, W. Wu, Z. Yu, and Q. Han, “Activecrowd: A framework for optimized multitask allocation in mobile crowdsensing systems,” *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 392–403, 2017.

[14] Y. Liu, B. Guo, Y. Wang, W. Wu, Z. Yu, and D. Zhang, “Taskme: multi-task allocation in mobile crowd sensing,” in *Proc. of ACM Ubicomp*, 2016, pp. 403–414.

[15] L. Wang, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, and Y. Wang, “Ccs-ta: Quality-guaranteed online task allocation in compressive crowdsensing,” in *Proc. of ACM UbiComp*, 2015, pp. 683–694.

[16] L. Wang, D. Zhang, Y. Wang, C. Chen, X. Han, and A. M’hamed, “Sparse mobile crowdsensing: challenges and opportunities,” *IEEE Communications Magazine*, vol. 54, no. 7, pp. 161–167, 2016.

[17] S. He, D.-H. Shin, J. Zhang, and J. Chen, “Toward optimal allocation of location dependent tasks in crowdsensing,” in *Proc. of IEEE INFOCOM*, 2014, pp. 745–753.

[18] Z. Wang, R. Tan, J. Hu, J. Zhao, Q. Wang, F. Xia, and X. Niu, “Heterogeneous incentive mechanism for time-sensitive and location-dependent crowdsensing networks with random arrivals,” *Computer Networks*, vol. 131, no. 2, pp. 96–108, 2018.

[19] H. Xiong, D. Zhang, G. Chen, L. Wang, and V. Gauthier, “Crowdtasker: Maximizing coverage quality in piggyback crowdsensing under budget constraint,” in *Proc. of IEEE PerCom*, 2015, pp. 55–62.

[20] D. Zhang, H. Xiong, L. Wang, and G. Chen, “Crowdrecruiter: selecting participants for piggyback crowdsensing under probabilistic coverage constraint,” in *Proc. of ACM UbiComp*, 2014, pp. 703–714.

[21] Z. Wang, J. Hu, J. Zhao, D. Yang, H. Chen, and Q. Wang, “Pay on-demand: Dynamic incentive and task selection for location-dependent mobile crowdsensing systems,” in *Proc. of IEEE ICDCS*, 2018.

[22] J. Lin, D. Yang, M. Li, J. Xu, and G. Xue, “Frameworks for privacy-preserving mobile crowdsensing incentive mechanisms,” *IEEE Transactions on Mobile Computing*, 2017.

[23] C. Dwork, “Differential privacy: A survey of results,” in *Proc. of TAMC*, 2008, pp. 1–19.

[24] G. Ghinita, “Privacy for location-based services,” *Morgan & Claypool Publishers*, vol. 4, no. 1, pp. 1–85, 2013.

[25] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, “Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy,” *IEEE Transactions on Dependable and Secure Computing*, 2016.

[26] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, “Geo-indistinguishability: Differential privacy for location-based systems,” in *Proc. of ACM CCS*, 2013, pp. 901–914.

[27] Y. Xiao and L. Xiong, “Protecting locations with differential privacy under temporal correlations,” in *Proc. of ACM CCS*, 2015, pp. 1298–1309.

[28] B. Wang and J. Yang, "Personalized (α, k)-anonymity algorithm based on entropy classification," *Journal of Computational Information Systems*, vol. 8, no. 1, pp. 259–266, 2012.

[29] X. Xiao and Y. Tao, "Personalized privacy preservation," in *Proc. of ACM SIGMOD*, 2006, pp. 229–240.

[30] X. Ye, Y. Zhang, and M. Liu, "A personalized (a, k)-anonymity model," in *Proc. of IEEE WAIM*, 2008, pp. 341–348.

[31] M. Alaggan, S. Gambs, and A.-M. Kermarrec, "Heterogeneous differential privacy," *Journal of Privacy and Confidentiality*, vol. 7, no. 2, pp. 127–158, 2016.

[32] Z. Jorgensen, T. Yu, and G. Cormode, "Conservative or liberal? personalized differential privacy," in *Proc. of IEEE ICDE*, 2015, pp. 1023–1034.

[33] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. of ACM MobiCom*, 2012, pp. 173–184.

[34] D. Yang, D. Zhang, V. W. Zheng, and Z. Yu, "Modeling user activity preference by leveraging user spatial temporal characteristics in lbsns," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 1, pp. 129–142, 2015.

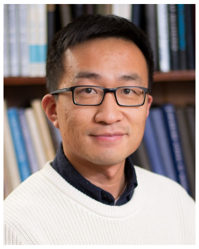
Zhibo Wang received the B.E. degree in Automation from Zhejiang University, China, in 2007, and his Ph.D degree in Electrical Engineering and Computer Science from University of Tennessee, Knoxville, in 2014. He is currently an Associate Professor with the School of Cyber Science and Engineering, Wuhan University, China. His currently research interests include mobile crowdsensing systems, cyber-physical systems, recommender systems and privacy protection. He is a Senior Member of IEEE and a Member of ACM.

Jiahui Hu received the B.S. degree in Information Security from Wuhan University, China, in 2016. She is currently pursuing her Master degree at School of Cyber Science and Engineering, Wuhan University. Her research interest focuses on mobile crowdsensing systems.

Ruizhao Lv received the B.S degree in Computing Science and Technology from Northeast Agricultural University, China, in 2017. He is currently pursuing his Master degree at School of Cyber Science and Engineering, Wuhan University. His research interest focuses on mobile crowdsensing systems.

Jian Wei received the B.S degree in Information and computing Science from Wuhan Textile University, China, in 2017. He is currently pursuing his Master degree at School of Computer, Wuhan University. His research interest focuses on mobile crowdsensing systems.

Qian Wang received the B.S. degree from Wuhan University, China, in 2003, the M.S. degree from Shanghai Institute of Microsystem and Information Technology, Chinese Academy of Sciences, China, in 2006, and the Ph.D. degree from Illinois Institute of Technology, USA, in 2012, all in Electrical Engineering. He is currently a Professor with the School of Cyber Science and Engineering, Wuhan University. His research interests include wireless network security and privacy, cloud computing security, and applied cryptography. Qian is an expert under "1000 Young Talents Program" of China. He is a co-recipient of the Best Paper Award from IEEE ICNP 2011. He is a Member of IEEE and a Member of ACM.



Dejun Yang received the B.S. degree from Peking University, Beijing, China, in 2007 and the Ph.D. degree in computer science from Arizona State University, Tempe, AZ, USA, in 2013. Currently, he is the Ben L. Fryrear Assistant Professor of computer science with Colorado School of Mines, Golden, CO, USA. His research interests include economic and optimization approaches to networks, crowdsourcing, smart grid, and security and privacy. Prof. Yang has served as a Technical Program Committee Member for many conferences, including the IEEE International Conference on Computer Communications (INFOCOM), the IEEE International Conference on Communications (ICC), and the IEEE Global Communications Conference (GLOBECOM). He has received Best Paper Awards at the IEEE GLOBECOM (2015), the IEEE International Conference on Mobile Ad hoc and Sensor Systems (2011), and the IEEE ICC (2011 and 2012), as well as a Best Paper Award Runner-up at the IEEE International Conference on Network Protocols (2010).

Hairong Qi received the B.S. and M.S. degrees in Computer Science from Northern JiaoTong University, Beijing, China in 1992 and 1995 respectively, and the Ph.D. degree in Computer Engineering from North Carolina State University, Raleigh, in 1999. She is currently the Gonzalez Family Professor with the Department of Electrical Engineering and Computer Science at the University of Tennessee, Knoxville. Her current research interests are in advanced imaging and collaborative processing in resource-constrained distributed environment, hyperspectral image analysis, and bioinformatics. She is a Fellow of IEEE.