Bilateral Privacy-Preserving Worker Selection in Spatial Crowdsourcing

Hengzhi Wang[1], Yongjian Yang[1], En Wang*[1], Xiulong Liu[2], Jingxiao Wei[1], Jie Wu[3], *Fellow, IEEE*

Abstract—Spatial Crowdsourcing (SC) has been adopted in various applications such as Gigwalk and Uber, where a platform takes location-based tasks (e.g., picking up passengers) from a requester and selects suitable workers to perform them. In most existing works, the platform selects workers based on the requester and worker information, which suffers from serious privacy issues. Some works have considered privacy issues, but they still suffer from either of two limitations: (i) Privacy of the requester and worker cannot be protected simultaneously; (ii) Third-party trusted entities are usually required. Motivated by this, we focus on protecting the privacy of both the requester and the worker without third-party entities while selecting workers. We use randomized response, a widely recognized and prevalent privacy model achieving Local Differential Privacy (LDP), to jointly protect the privacy of workers' locations and charges based on the location-charge correlation. For the requester, we present a novel mechanism called randomized matrix multiplication to hide the real task locations. More importantly, we prove that the worker selection based on the protected information is non-submodular and NP-hard, which cannot be addressed in polynomial time. To this end, we present an approximate algorithm to solve the problem efficiently, of which the effectiveness is measured by the approximation ratio, i.e., the ratio of the optimal solution to the approximate solution. Finally, simulations based on real-world datasets illustrate that our worker selection outperforms the state-of-the-art method on both privacy protection and worker selection.

Index Terms—Spatial crowdsourcing, bilateral privacy-preserving, worker selection, local differential privacy.

1 Introduction

With the rapid development of mobile Internet and devices, spatial crowdsourcing, a novel paradigm of exploring the power of crowd to collect and share data, has emerged in recent years [1]–[3]. In typical scenarios, e.g., worker selection [4], [5] and task allocation [6], [7], requesters and workers need to report their real information to the spatial crowdsourcing platform. For example, some Gigwalk requesters have to upload their task locations to the platform, and Gigwalk workers need to upload their physical locations and charges for each task as well. Since the spatial crowdsourcing platform may not be trustworthy, several privacy concerns have arisen [8], e.g., the information of tasks and workers is leaked or used illegally by the platform, which violates the General Data Protection Regulation (GDPR) [9].

To this end, many privacy-preserving worker selection strategies [10]–[15] are proposed to address the privacy issues by adding noise (e.g., Laplace mechanism [10], [11]) to the item or obfuscating the item to another one (e.g., obfuscation function [12]). We argue that these strategies only focus on protecting the privacy of workers' locations [10], [12], or the privacy of workers' charges [13] while ignoring the privacy of tasks' locations. Actually, workers' locations, charges, and tasks' locations are also critical parts of the sensitive information that should be protected as well, and the leakages on them to the untrusted platform would discourage participation in spatial crowdsourcing activities. Several works [14], [15] aim to protect the privacy of both tasks and workers, which unfortunately have to rely on third-party trusted entities, resulting in non-negligible service cost and communication overhead.

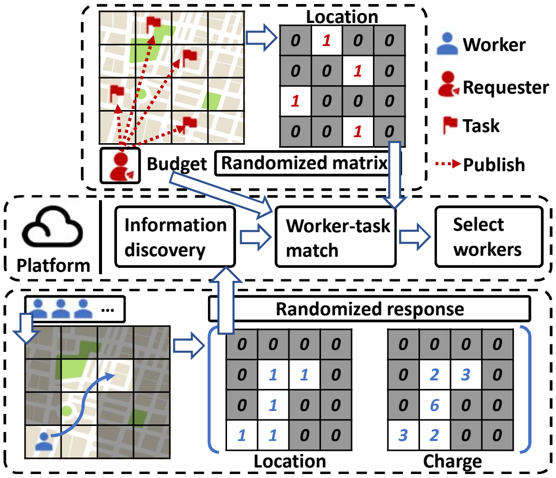


Fig. 1: The bilateral privacy-preserving worker selection framework without relying on third-party trusted entities.

Motivated by these privacy issues, this paper focuses on designing a *bilateral* privacy-preserving worker selection strategy that could select suitable workers as well as protect the privacy of both requesters and workers without third-

- [1]Hengzhi Wang, Yongjian Yang, En Wang and Jingxiao Wei are with the Department of Computer Science and Technology and Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education, Jilin University, Changchun, Jilin 130012, China. (E-mails: wanghz17@mails.jlu.edu.cn; yyj@jlu.edu.cn; wangen@jlu.edu.cn; weijx19@mails.jlu.edu.cn)
- [2]Xiulong Liu is with the College of Intelligence and Computing, Tianjin University, China. (E-mail: xiulong_liu@tju.edu.cn)
- [3]Jie Wu is with the School of Computer and Information Sciences, Temple University, USA. (E-mail: jiewu@temple.edu)

*The corresponding author is En Wang.

party trusted entities. A specific example is shown in Fig. 1. The platform receives the obfuscated information reported from the requester and workers, and selects appropriate workers under the budget constraint to maximize the total task completion effect. Specifically, the spatial area is divided into multiple subareas. For the requester (the upper part of Fig. 1), it publishes a set of tasks in different subareas, i.e., each task contains a specific location. Thus, the privacy of the requester refers to the privacy of the tasks' locations. Then, the requester hides tasks' real locations and provides the obfuscated locations and a total payment budget for these tasks to the platform. For workers (the lower part of Fig. 1), each one contains the following information: (i) future locations that workers will cover before the task deadline; (ii) the corresponding charges of performing tasks at particular locations. Then, the requester and the workers obfuscate their original information and report the obfuscated one to the platform. Based on this, the platform selects workers under the payment budget constraint. The idea of this paper is that we not only obfuscate the sensitive information against the platform, but also achieve a good worker selection performance based on the obfuscated information.

Selecting workers with sanitized information from the workers and the requester is more difficult. Moreover, the third-party involvement is not allowed as well. Thus, the first challenge is how to *construct a bilateral privacy-preserving framework* to protect the privacy of the requester and the workers. To address the challenge, two privacy models and methods are adopted respectively taking into account the characterizations of large-scale workers and a single requestor. More precisely, following Local Differential Privacy (LDP) [16], a widely recognized privacy model with distributed architecture requiring a large number of individuals, we advise an Extended Randomized Response (ERR) based on [17], [18] for workers to obfuscate their real locations and corresponding charges. This ensures that each individual worker's sensitive information is protected from privacy leakage on the untrusted platform while the statistics of the large-scale workers can be well estimated and used. For the requester, since there is only a single requester in the system, LDP and ERR, unfortunately, cannot be adopted to deal with the requester's privacy. To this end, we present a Randomized Matrix Multiplication (RMM) based on [14] to obfuscate the requestor's task locations and use ϵ-Privacy following the location privacy model [19] to measure the requestor's privacy. Both ERR and RMM protect privacy locally and do not rely on third-party trusted.

Furthermore, when protecting workers' privacy, we consider protecting both the locations (categorical value) and charges (numerical value) of workers. An easy solution is to protect them separately. However, this solution could be improved because it fails to utilize the potential correlation between the location and the charge. In other words, the location confusion result could be regarded as the prior knowledge, with the help of which, workers could achieve more efficient charge confusion and the platform can also recover the information of a worker set accurately. Hence, how to *utilize workers' location-charge correlation* to achieve more efficient privacy protection of workers is the second challenge. To address the second challenge, we first use RR to obfuscate workers' locations, and then conditional RR is used to achieve a more efficient charge confusion based on the confusion results of workers' locations. In this way, the possible passing locations and the estimated total charge for the selected workers could be more accurately recovered. Moreover, under such a bilateral privacy-preserving framework, all the collected information from the requester and worker is obfuscated. In this case, the worker selection problem is proven to be non-submodular and non-monotone. Thus, how to *select suitable workers based on the obfuscated information* is the third challenge. To handle the third challenge, we prove that the privacy-preserving worker selection problem is NP-hard and present an approximate algorithm inspired by the constrained non-monotone non-submodular maximization problem. Through rigorous proof, we verify that the algorithm achieves a lower bound. Then, our main contributions are summarized as follows:

- We present a bilateral privacy-preserving framework to select suitable workers under the budget constraint while protecting the privacy of both the requester and the worker, without relying on third-party trusted entities.
- To protect the privacy of workers, we propose a local confusion strategy based on extended randomized response to achieve local differential privacy, which utilizes the location-charge correlation to protect both workers' locations and charges.
- We prove that the bilateral privacy-preserving worker selection problem is NP-hard, non-submodular and non-monotone, and propose an approximate algorithm based on the greedy method to achieve a lower bound $\gamma{\cdot}(1{-}1/e)$ of the worker selection performance.
- We conduct extensive simulations based on three real-world datasets to evaluate the performance of the proposed algorithm, and the results show that our algorithm always outperforms the state-of-the-art strategy.

The remainder of the paper is organized as follows. We review the related works in Section II. The system model and the formulated problem are described in Section III. The bilateral privacy-preserving framework is proposed in Section IV. In Section V, we analyze the performance of the proposed algorithm. Finally, we conduct simulations in Section VI and conclude the paper in Section VII.

2 RELATED WORK

In this section, we introduce the existing works on privacy-preserving worker selection, privacy-preserving crowdsourced data analysis, and local differential privacy.

Privacy-preserving Worker Selection: Wang *et al.* [13] considered that workers' bids (i.e., utilities) submitted to the platform were the sensitive information of workers. To this end, they proposed a bid obfuscation function against the untrusted platform based on the exponentiation mechanism. Prorok *et al.* [20] presented privacy-preserving worker selection strategies based on the Hungarian algorithm by adding random noise drawn from a two-dimensional Laplace distribution. Xiao *et al.* [21] considered the privacy of the worker information against the platform, and proposed a secure worker selection protocol based on the semi-honest model. However, all of them [13], [20], [21] ignored the privacy of the requester. Furthermore, Shahabi *et al.* [10] and Zhao *et al.* [15] considered the privacy of both requesters and workers, then proposed the bilateral privacy-preserving task allocation strategy by adding random noise [10] or Paillier cryptosystem [15], respectively. But both of them only considered workers' locations and ignored workers' charges/costs. In addition, Paillier cryptosystem also needed a third-part trusted entity to distribute the public key. Hence, the existing privacy-preserving strategies cannot be directly used to address the problem in this paper.

Privacy-preserving Data Analysis: There have been many works focusing on privacy-preserving data analysis. Taking privacy concerns in crowdsourcing data publication into consideration, Ren *et al.* [22] designed a data publication method using estimation maximization for achieving locally differential privacy. In their design, data correlations are identified to promote data utility. A similar problem was studied by Wang *et al.* [23], in which high-dimensional crowdsourcing data is synthesized via attribute dependence and is released in a randomized response way. Consequently, the trade-off between privacy-preserving data releasing and data utility is balanced. Apart from data releasing, the privacy-preserving statistic estimation was studied in [24] using spatial-temporal correlations in real-time crowdsourcing data, and the privacy-preserving truthful reference was investigated in [25] given sparse worker answers. Yet, the aforementioned privacy-preserving strategies are designated for data analysis that, unfortunately, cannot be employed for worker selection in this paper.

Local Differential Privacy: Differential privacy was first proposed by Dwork [26], which mainly includes centralized differential privacy [27], [28] and local differential privacy [16], [29]. Centralized differential privacy assumes a trusted central data collector to possess data, which brings the non-negligible service charge. Thus, many existing works focused on Local Differential Privacy (LDP). In LDP, Erlingsson *et al.* [30] first considered the frequency estimation over categorical data based on randomized response (RR). Then, Fanti *et al.* [31] extended [30] to a more practical scenario, which could estimate unknown strings to learn without explicit knowledge. However, [30] and [31] only applied LDP for categorical data. To this end, Ye *et al.* [16] considered the estimations over both categorical and numerical data, and proposed PrivKV based on LDP. Yet, PrivKV only focused on a fixed worker set. In this paper, the worker set is dynamically changing depending on the worker selection method.

3 System model and Problem formulation

In this section, we introduce the system model and privacy models and formulate the bilateral privacy-preserving worker selection problem to be addressed.

3.1 System Model

The entire area is divided into $k \times k$ grids [32], which is denoted as a location matrix $\mathcal{L} = \{\ell_{xy} = (x, y) | 1 \leq x \leq k, 1 \leq y \leq k\}$, where the location ℓ_{xy} refers to the grid (x, y). In the following, we simplify ℓ_{xy} as ℓ, and the same simplification is applied for other notations. The requester publishes a set of tasks distributed in different subareas. Without loss of generality, we assume that each task corresponds to a location [5], [6], multiple tasks in a location can be treated as a complex task. In addition, we also assume that all tasks with the same deadline only need to be completed once. Next, the published task set can be denoted by a task location matrix $\mathcal{L}_j = \{\ell^j_{xy} | 1 \leq x \leq k, 1 \leq y \leq k\}$. We simplify ℓ^j_{xy} as ℓ_j, and for each item $\ell_j \in \mathcal{L}_j$, if there is a task in the location ℓ, then $\ell_j = 1$, otherwise $\ell_j = 0$.

TABLE 1: List of key notations

Notation	Description
$\mathcal{L}, U$	The sets of locations and total workers
$\mathcal{L}_j, \mathcal{L}^*_j$	The task location matrix, the obfuscated task location matrix
$\mathcal{L}_i, \mathcal{C}_i$	The location matrix and charge matrix of the worker u_i
$\mathcal{L}^*_i, \mathcal{C}^*_i$	The obfuscated location matrix and charge matrix of the worker u_i
$S, \mathcal{B}$	The set of selected workers from U, the budget of the total charge
ϵ_1, ϵ_2	The privacy budgets for the worker location privacy and worker charge privacy
p_1, p_2	The confusion probabilities for the worker location and worker charge
c_{min}, c_{max}	The minimum and maximum charge among all workers
$f_\ell(S, \epsilon_1)$	The number of times workers in S cover the location ℓ based on the obfuscated locations
$f^*_\ell(S, \epsilon_1)$	The estimated number of times workers in S cover the location ℓ based on the real locations
f_ℓ, f^*_ℓ	The simplifications of $f_\ell(S, \epsilon_1)$ and $f^*_\ell(S, \epsilon_1)$
$\mathcal{C}(S, \epsilon_2)$	The estimated total charge of all workers in S
$\mathcal{C}(S)$	The simplification of $\mathcal{C}(S, \epsilon_2)$
$F(S, \epsilon_1)$	The utility of all workers in S
$\mathcal{R}$	The invertible matrix generated by the requester
$\mathcal{L}(S)$	The location matrix of S, where each item denotes the estimated number of S covers each location in $\mathcal{L}$
U_t, S_t	The total worker set and selected worker set in the t-iteration when selecting workers
$c(S), c^*(S)$	The real total charge and estimated total charge of workers in S for the specific location ℓ
γ	The ratio measuring the submodularity of a non-submodular set function

The set of total workers is denoted as $U = \{u_1, u_2, \cdots, u_n\}$, where each worker u_i has a location matrix $\mathcal{L}_i = \{\ell^i_{xy} | 1 \leq x \leq k, 1 \leq y \leq k\}$ indicating that the locations u_i will cover before the task deadline. We simplify ℓ^i_{xy} as ℓ_i. For $\ell_i \in \mathcal{L}_i$, if u_i will cover the location ℓ, then $\ell_i = 1$, otherwise $\ell_i = 0$. As long as the worker covers the location of a task, the task could be completed [21], [33]. In addition, each worker u_i also has a charge matrix $\mathcal{C}_i = \{c^i_{xy} | 1 \leq x \leq k, 1 \leq y \leq k\}$, which denotes the charges of u_i performing tasks in the corresponding locations in $\mathcal{L}_i$. Hence, each worker u_i has two attribute matrices $(\mathcal{L}_i, \mathcal{C}_i)$. Intuitively, the charge matrix $\mathcal{C}_i$ is related to the location matrix $\mathcal{L}_i$, e.g., if a worker will not cover a location, its charge of that location is 0. That is, for $c_i \in \mathcal{C}_i$, if $\ell_i = 1$, then $c_i > 0$, otherwise $c_i = 0$. The description of key notations is presented in Table 1.

3.2 Privacy Model

In this paper, we assume that the platform is *honest-but-curious* [34], [35], which means that the platform will honestly carry out the task publishment and worker selection processes, but will also be curious about the sensitive information of the requester and worker. In addition, there are some differences between the requester and worker: (i)

There is only one requester, but there are multiple workers when reporting information; (ii) The requester only has one task location matrix, but each worker has a worker location matrix and a charge matrix. Therefore, we should construct the privacy models for the requester and workers respectively.

1) The requester privacy: Since the requester publishes a set of tasks represented by the task location matrix $\mathcal{L}_j$, the requester privacy is actually the task location privacy, i.e., the task location matrix $\mathcal{L}_j$ should be obfuscated to $\mathcal{L}_j^*$ to prevent the platform from violating the task location information.

2) The worker privacy: Since each worker in the worker set U not only has a location matrix $\mathcal{L}_i$ but also has a charge matrix $\mathcal{C}_i$, the worker privacy consists of two aspects: (i) the worker location privacy, i.e., the location matrix $\mathcal{L}_i$ should be obfuscated to $\mathcal{L}_i^*$; (ii) the worker charge privacy, i.e., the charge matrix $\mathcal{C}_i$ should be obfuscated to $\mathcal{C}_i^*$ to prevent the platform from violating the task location information.

To measure the privacy protection level, we define privacy budget in Definition 1.

Definition 1 (Privacy budget)**.** *The privacy budget $\epsilon \geq 0$ represents the privacy level that the mechanism provides. A smaller ϵ guarantees a stronger privacy level.*

Taking into account the characterizations of large-scale workers and a single requester, we define different privacy models for them. First, given a privacy budget ϵ, Local Differential Privacy (ϵ-LDP), a prevalent privacy model with distributed architecture, is offered in Definition 2 as in previous studies [16] to measure the workers' privacy. Second, following the location privacy model [19], we define ϵ-Privacy in Definition 3 to measure the requester's privacy.

Definition 2 (ϵ-LDP)**.** *A confusion strategy $\mathcal{M}$ satisfies ϵ-LDP when the probabilities of getting the same output for two inputs always meet the following inequality:*

$$\Pr[\mathcal{M}(d_1)=d^*]\leq e^{\epsilon}\cdot\Pr[\mathcal{M}(d_2)=d^*], \tag{1}$$

where d_1, d_2, d^* are the numerical values or matrices.

Definition 3 (ϵ-Privacy)**.** *A confusion strategy $\mathcal{M}$ satisfies ϵ-Privacy when the probabilities of the platform inferring a task's real data meet the following inequality:*

$$\Pr[d|d^*]\leq e^{\epsilon}\cdot\Pr[d], \tag{2}$$

where $\Pr[d]$ denotes the probability of inferring the real data d without any information, $\Pr[d|d^*]$ denotes the probability of inferring the real data given the obfuscated data d^*.

For the requester, since the item in the task location matrix of the requester is the categorical value, the traditional Laplace noise mechanism cannot be adopted to protect the requester privacy. In addition, there is only a single requester in this paper, so neither the obfuscation function [12] nor RR [17] can be applied since the statistic cannot be estimated accurately with small sample size. Thus, we present a novel strategy RMM in Section IV to obfuscate $\mathcal{L}_j$ to $\mathcal{L}_j^*$ while satisfying ϵ-Privacy.

For the worker, RR is a well-known technique to satisfy ϵ-LDP and suitable for multiple workers. However, in this paper, the worker retains both locations (categorical value) and charges (numerical value), thus we extend the traditional RR to obfuscate multiple information of workers. For worker locations, RR enables each worker to give a random answer to a boolean question, e.g., whether the worker will cover location ℓ. More precisely, for each question, the worker will tell a truth with probability p_1 or a lie with probability $1-p_1$. This way a real answer ℓ_i is transformed into an obfuscated answer ℓ_i^*. To satisfy ϵ-LDP, we set $p_1 = e^{\epsilon_1}/(1+e^{\epsilon_1})$. After the platform obtains the obfuscated answers to the question from a set of workers S, i.e., the subset of U, let $f_\ell(S,\epsilon_1)$ denote the count of "Yes" in all obfuscated answers and N the count of all obfuscated answers. The platform estimates the count of "Yes" in real answers $f_\ell^*(S,\epsilon_1)$ based on the count of "Yes" in $f_\ell(S,\epsilon_1)$:

$$f_\ell^*(S,\epsilon_1)=(p_1-1)\cdot N/(2p_1-1)+f_\ell(S,\epsilon_1)/(2p_1-1), \tag{3}$$

where $N=|S|$. In the following sections, we will prove that $f_\ell^*(S,\epsilon_1)$ is an unbiased estimation of $f_\ell(S,\epsilon_1)$. Accordingly, in our problem, $\ell_i^*=1$ means that u_i gives answer "Yes" to the question whether the worker will cover location ℓ. Hence, the estimated number of times workers in S cover location ℓ can be denoted by $f_\ell^*(S,\epsilon_1)$ in Eq. 3.

Similarly, for worker charges, different from the location ℓ_i, the worker's charge c_i is not a discrete categorical value, but a continuous numerical value $c_i \in [c_{min}, c_{max}]$, where c_{min} and c_{max} are the minimum and maximum charges among all worker charges. Considering that, we first discretize the charge of each worker c_i into c_{min} or c_{max} locally. Furthermore, based on RR and the location-charge correlation, each worker gives an obfuscated answer to the question whether the worker's charge is c_{max} for covering the location ℓ. In our problem, $c_i^* = c_{max}$ means that u_i gives answer "Yes", and $c_i^* = c_{min}$ for "No". Let $n_1(S,\epsilon_2)$, $n_2(S,\epsilon_2)$ denote the counts of c_{max} and c_{min} in all obfuscated answers of S, respectively. Based on these, the platform estimates the counts of c_{max} and c_{min} in real answers $n_1^*(S,\epsilon_2)$, $n_2^*(S,\epsilon_2)$ similar to $f_\ell^*(S,\epsilon_1)$ in Eq. 3. Note that $N = n_1(S,\epsilon_2)+n_2(S,\epsilon_2)$. Then, the estimated total charge $\mathcal{C}(S,\epsilon_2)$ of S is denoted as follows:

$$\mathcal{C}(S,\epsilon_2)=\sum\nolimits_{\ell\in\mathcal{L}}[n_1^*(S,\epsilon_2)\cdot c_{max}+n_2^*(S,\epsilon_2)\cdot c_{min}]. \tag{4}$$

Based on Eqs. 3-4, even if a set of workers S upload the obfuscated locations and charges, the platform can still estimate the useful information of S, i.e., $f_\ell^*(S,\epsilon_1)$ and $\mathcal{C}(S,\epsilon_2)$. Then, the platform measures the utility of a set of workers based on $f_\ell^*(S,\epsilon_1)$ and $\mathcal{L}_j^*$ as shown in Definition 4.

Definition 4 (The worker's utility)**.** *The utility of a set of workers S is defined as the expected number of tasks covered by the workers in S,*

$$F(S,\epsilon_1)=\sum\nolimits_{\ell\in\mathcal{L}}\min\{f_\ell^*(S,\epsilon_1),1\}\cdot x_\ell, \tag{5}$$

where 1 indicates that each task only needs to be completed once, $x_\ell \in \{0,1\}$ denotes whether there is a task in location ℓ and is calculated based on $f_\ell^*(S,\epsilon_1)$ and $\mathcal{L}_j^*$.

3.3 Problem Formulation

Our problem has emerged as follows:

Problem [Bilateral privacy-preserving worker selection under the budget constraint]: In the area $\mathcal{L}$, given the task set $\mathcal{L}_j$ and worker set U, the honest-but-curious platform selects a set of workers S based on the obfuscated information

Algorithm 1 Task Location Confusion (TLC)

```
Input: Task location matrix $\mathcal{L}_j$
Output: Collection of obfuscated task location matrices $\Gamma$
1: $\Gamma \leftarrow \emptyset$
2: Generate a $k \times k$ invertible real number matrix $\mathcal{R}$
3: for each $\ell_j \in \mathcal{L}_j$ and $\ell_j = 1$ do
4:     $\mathcal{L}'_j \leftarrow \{0, \cdots, 0, \ell_j, 0, \cdots, 0\}$
5:     $\mathcal{L}^*_j \leftarrow (\mathcal{L}'_j)^T \times \mathcal{R}^T \times \mathcal{R}$
6:     $\Gamma \leftarrow \Gamma \cup \{\mathcal{L}^*_j\}$
7: return $\Gamma$
```

of tasks and workers under the payment budget constraint. The objective is to maximize the expected number of completed tasks:

$$\begin{aligned} &\text{Maximize} \quad F(S,\epsilon_1)=\sum\nolimits_{\ell\in\mathcal{L}} \min\{f^*_\ell(S,\epsilon_1),1\}\cdot x_\ell, \\ &\text{Subject to} \quad S\subseteq U,\ \ \mathcal{C}(S,\epsilon_2)\leq\mathcal{B}. \end{aligned} \tag{6}$$

The first constraint indicates that workers in S are selected from the whole worker set U, and the second constraint indicates that we cannot select workers limitlessly, i.e., the expected total charge $\mathcal{C}(S,\epsilon_2)$ of S should be limited by $\mathcal{B}$. $\mathcal{B}$ is the budget of the total charge.

4 Bilateral Privacy-preserving Strategy Design

4.1 The Confusion Strategy for The Requester

Considering the requester privacy, the requester should obfuscate the task location matrix $\mathcal{L}_j$ to $\mathcal{L}^*_j$ before uploading it to the platform. To this end, we extend the matrix multiplication [14] and present a novel strategy RMM. Based on RMM, we propose a Task Location Confusion strategy (TLC) in Algorithm 1 to hide the real information, which is performed by the requester locally.

The requester randomly generates a $k \times k$ invertible real number matrix $\mathcal{R}$ (line 2). For each task satisfying $\ell_j \in \mathcal{L}_j$ and $\ell_j = 1$, we reconstruct a task location matrix $\mathcal{L}'_j$, where all items are equal to 0 apart from ℓ_j (line 4). We calculate the obfuscated task location matrix $\mathcal{L}^*_j$ and add it to the collection of matrices Γ (line 5), where $(\mathcal{L}'_j)^T$, $\mathcal{R}^T$ represent the transposes of $\mathcal{L}'_j$, $\mathcal{R}$. An example is shown in Fig. 2, and we discover that the task location matrix $\mathcal{L}_j$ is decomposed into multiple sparse matrices. Then, each sparse matrix is obfuscated before uploading to the platform. Based on the obfuscated matrices, the platform can only know which column the task is in rather than its specific location, thus the privacy of tasks' locations is protected. In fact, the platform matches the obfuscated task location matrix with the estimated worker location matrix without knowing the real task location to identify whether a task can be covered, thereby determining the utility of a set of workers. Next, we prove that TLC satisfies ϵ-Privacy in Theorem 1.

Theorem 1. *TLC achieves ϵ-Privacy, where $\epsilon = \ln k$.*

Proof. Given a reconstructed matrix $\mathcal{L}'_j$ and an obfuscated matrix $\mathcal{L}^*_j$ as shown in lines 4-5 of Algorithm 1, let $\Pr[\mathcal{L}'_j]$ denote the probability of the platform inferring the real matrix $\mathcal{L}'_j$ without any extra information and $\Pr[\mathcal{L}'_j|\mathcal{L}^*_j]$ denote the probability of inferring $\mathcal{L}'_j$ with knowing $\mathcal{L}^*_j$. Thus, we have $\Pr[\mathcal{L}'_j] = 1/k^2$ and $\Pr[\mathcal{L}'_j|\mathcal{L}^*_j] = 1/k$. According to Definition 3, we obtain $\Pr[\mathcal{L}'_j|\mathcal{L}^*_j]/\Pr[\mathcal{L}'_j] = k \leq e^{\ln k}$, which means the knowledge brought by the obfuscated information $\mathcal{L}^*_j$ is limited and bounded. Therefore, TLC achieves ϵ-Privacy, where $\epsilon = \ln k$. □

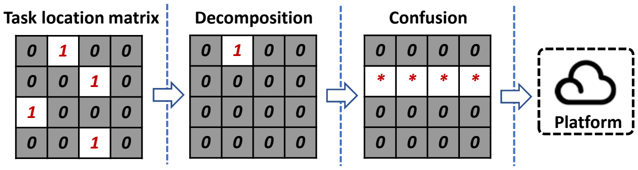


Fig. 2: An example for TLC.

Algorithm 2 Worker Charge Confusion (WCC)

```
Input: Charge $c$, privacy budget $\epsilon$
Output: Obfuscated charge $c^*$
1: Discretize $c$ based on Eq. 7
2: $c^* \leftarrow c$
3: Generate a probability $p \in [0,1]$ randomly
4: If $p \leq \frac{1}{1+e^\epsilon}$, then $c^* \leftarrow c_{max} + c_{min} - c$
5: return $c^*$
```

4.2 The Confusion Strategy for Workers

Based on RR, we propose the obfuscated strategy for workers. As mentioned above, each worker u_i has a location matrix $\mathcal{L}_i$ and a charge matrix $\mathcal{C}_i$. For a specific location ℓ, the worker u_i actually holds a location-charge pair $[\ell_i, c_i]$, where ℓ_i is a categorical value (1 or 0), and $c_i \in [c_{min}, c_{max}]$ is a continuous value. Hence, to obfuscate the location and charge together, we first discretize each worker's charge:

$$c_i \leftarrow \begin{cases} c_{max}, & \text{w.p. } \frac{c_i - c_{min}}{c_{max} - c_{min}}. \\ c_{min}, & \text{w.p. } \frac{c_{max} - c_i}{c_{max} - c_{min}}. \end{cases} \tag{7}$$

Then, we propose the Worker Charge Confusion strategy (WCC) depicted in Algorithm 2. We input a charge c and the privacy budget ϵ. Then, c is unchanged with probability $e^\epsilon/(1+e^\epsilon)$ or obfuscated with probability $1/(1+e^\epsilon)$ (line 4).

After the charge discretization, we obfuscate the worker's location ℓ_i by applying RR. Note that there is a correlation between the worker's location and charge for each location. Specifically, for a location-charge pair $[\ell_i, c_i]$, if $\ell_i = 0$, the corresponding charge should also be 0 because the charge cannot be greater than 0 when the worker does not cover ℓ_i. With these in mind, let $1 \rightarrow 1$ denote that $\ell_i = 1$ is obfuscated to $\ell^*_i = 1$, then we obtain four cases:

1) $1 \rightarrow 1$: In this case, since the location has not changed, the location-charge pair can remain unchanged after confusion, i.e., $[1, c_i] \rightarrow [1, c_i]$;
2) $1 \rightarrow 0$: In this case, since the location has changed, the charge should be 0 to protect the worker's real location, i.e., $[1, c_i] \rightarrow [0, 0]$;
3) $0 \rightarrow 1$: Since the location has changed, a new location-charge pair needs to be generated. Considering the location-charge correlation, we assign the mean of charges to the new pair because the real charge should be greater than 0 when the worker's location $\ell_i = 1$, i.e., $[0, 0] \rightarrow [1, (c_{max} + c_{min})/2]$;
4) $0 \rightarrow 0$: In this case, the location has not changed, and the worker does not cover ℓ, so the charge should be 0, i.e., $[0, 0] \rightarrow [0, 0]$.

Algorithm 3 Worker Location Confusion (WLC)

```
Input: A location-charge pair $[\ell_i, c_i]$, privacy budgets $\epsilon_1$, $\epsilon_2$
Output: Obfuscated location-charge pair $[\ell_i^*, c_i^*]$
 1: if $\ell_i = 1$ then
 2:     $c_i^* = \mathrm{WCC}(c_i,\ \epsilon_2)$
 3:     Generate a probability $p \in [0,1]$ randomly
 4:     if $p \le 1/(1+e^{\epsilon_1})$, then $[\ell_i^*, c_i^*] \leftarrow [0,0]$
 5:     if $p > 1/(1+e^{\epsilon_1})$, then $[\ell_i^*, c_i^*] \leftarrow [1, c_i^*]$
 6: else
 7:     $c_i^* = \mathrm{WCC}((c_{max}+c_{min})/2,\ \epsilon_2)$
 8:     Generate a probability $p \in [0,1]$ randomly
 9:     if $p \le 1/(1+e^{\epsilon_1})$, then $[\ell_i^*, c_i^*] \leftarrow [1, c_i^*]$
10:     if $p > 1/(1+e^{\epsilon_1})$, then $[\ell_i^*, c_i^*] \leftarrow [0,0]$
11: return $[\ell_i^*, c_i^*]$
```

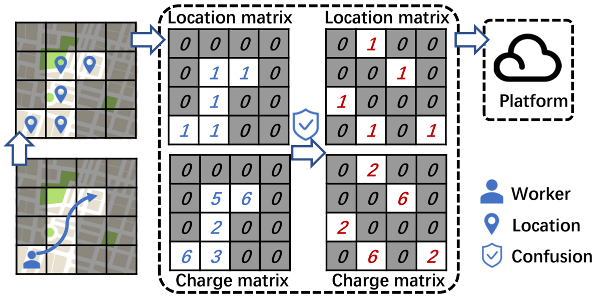


Fig. 3: Worker location confusion.

We then present the Worker Location Confusion (WLC) strategy in Algorithm 3 to obfuscate the location-charge pair according to [16]. The inputs are a location-charge pair and privacy budgets, where the privacy budget ϵ_1 is used for confusing the location and ϵ_2 is used for the charge. Considering the four cases, WLC first determines whether the worker's real location ℓ_i is 1 (lines 1 and 6), then obfuscates the worker's charge based on WCC, and gets an obfuscated charge c_i^* (lines 2 and 7) . Then, the location-charge pair is obfuscated according to the four cases (lines 4-5 and 9-10). An example of WLC is shown in Fig. 3. Specifically, according to the real trajectory, each worker generates the location matrix and charge matrix for each location $\ell \in \mathcal{L}$. Then, for privacy protection, each worker obfuscates the matrices locally in the form of location-charge pairs based on the worker location confusion strategy proposed in Algorithm 3 and uploads them to the platform. Next, we prove that WLC satisfies ϵ-LDP in Theorem 2.

Theorem 2. *WLC achieves ϵ-LDP, where $\epsilon = \epsilon_1+\epsilon_2$, ϵ_1 and ϵ_2 are the privacy budgets for locations and charges.*

Proof. Given the worker u_i and location ℓ, let $[\ell_i^*, c_i^*]$ be the obfuscated location-charge pair output by WLC (Algorithm 3). First, we focus on the location confusion. For the case of the obfuscated location $\ell_i^* = 1$, assuming that there are two distinct real locations ℓ_i^1 and ℓ_i^2, according to WLC, we have

$$\frac{\Pr[\ell_i^*|\ell_i^1]}{\Pr[\ell_i^*|\ell_i^2]} \le \frac{\Pr[\ell_i^*=1|\ell_i=1]}{\Pr[\ell_i^*=1|\ell_i=0]} = (\frac{e^{\epsilon_1}}{1+e^{\epsilon_1}})/(\frac{1}{1+e^{\epsilon_1}}) = e^{\epsilon_1}, \quad (8)$$

where $\Pr[\ell_i^*|\ell_i^1]$ indicates the probability of ℓ_i^1 being obfuscated to ℓ_i^*. The case of $\ell_i^* = 0$ can be analyzed in the similar manner. Thus, the confusion on locations in WLC satisfies ϵ_1-LDP according to Definition 2. Next, for the charge confusion, considering the location-charge correlation, we investigate two cases: $[\ell_i^*, c_i^*] = [0,0]$ or $[1, c_i^*]$. When $[\ell_i^*, c_i^*] = [0,0]$, assuming that there are two distinct real charges c_i^1 and c_i^2, we have

$$\frac{\Pr[c_i^*|c_i^1]}{Pr[c_i^*|c_i^2]} \le \frac{\Pr[c_i^*=0|c_i^1=c_{max}]\cdot\Pr[\ell_i^*=0]}{Pr[c_i^*=0|c_i^2=c_{min}]\cdot\Pr[\ell_i^*=0]}$$
$$= (\frac{e^{\epsilon_2}}{1+e^{\epsilon_2}})/(\frac{1}{1+e^{\epsilon_2}}) = e^{\epsilon_2}. \quad (9)$$

When $[\ell_i^*, c_i^*]$=$[1, c_i^*]$, we focus on the case c_i^*=c_{max} and get

$$\frac{\Pr[c_i^*|c_i^1]}{\Pr[c_i^*|c_i^2]} = \frac{c_i^1\cdot\frac{e^{\epsilon_2}}{1+e^{\epsilon_2}} + (c_{max}+c_{min}-c_i^1)\cdot\frac{1}{1+e^{\epsilon_2}}}{c_i^2\cdot\frac{e^{\epsilon_2}}{1+e^{\epsilon_2}} + (c_{max}+c_{min}-c_i^2)\cdot\frac{1}{1+e^{\epsilon_2}}}$$
$$= \frac{(e^{\epsilon_2}-1)\cdot c_i^1 + c_{max}+c_{min}}{(e^{\epsilon_2}-1)\cdot c_i^2 + c_{max}+c_{min}}$$
$$\le \frac{e^{\epsilon_2}\cdot c_{max}+c_{min}}{e^{\epsilon_2}\cdot c_{min}+c_{max}} \le e^{\epsilon_2}. \quad (10)$$

The first equality holds because $c_{min} \le c_i^1, c_i^2 \le c_{max}$. Then, the case $c_i^* = c_{min}$ can be analyzed in the similar manner. Thus, the confusion on charges in WLC satisfies ϵ_2-LDP according to Definition 2. Moreover, according to the sequential composition [16], WLC satisfies $(\epsilon_1$+$\epsilon_2)$-LDP. □

Algorithm 4 Platform Calibration Process (PCP)

```
Input: For each worker $u_i \in S$, obfuscated location matrix $\mathcal{L}_i^*$, obfuscated charge matrix $\mathcal{C}_i^*$, privacy budgets $\epsilon_1$ and $\epsilon_2$
Output: Location matrix $\mathcal{L}(S)$, total charge $\mathcal{C}(S)$
 1: $\mathcal{L}(S) \leftarrow \emptyset, \mathcal{C}(S) \leftarrow 0$
 2: $p_1 \leftarrow e^{\epsilon_1}/(e^{\epsilon_1}+1)$, $p_2 \leftarrow e^{\epsilon_2}/(e^{\epsilon_2}+1)$
 3: for each $\ell \in \mathcal{L}$ do
 4:     $f_\ell \leftarrow$ Count$(\ell_i^* = 1)$, $\forall u_i \in S$
 5:     $n_1 \leftarrow$ Count$(c_i^* = c_{max})$, $\forall u_i \in S$
 6:     $n_2 \leftarrow$ Count$(c_i^* = c_{min})$, $\forall u_i \in S$
 7:     $N = n_1 + n_2$
 8:     Calibrate $f_\ell^*$ as: $f_\ell^* \leftarrow \frac{(p_1-1)\cdot|S|}{2p_1-1} + \frac{f_\ell}{2p_1-1}$
 9:     Calibrate $n_1^*, n_2^*$ as:
10:     $n_1^* \leftarrow \frac{p_2-1}{2p_2-1}\cdot N + \frac{n_1}{2p_2-1}$, $n_2^* \leftarrow \frac{p_2-1}{2p_2-1}\cdot N + \frac{n_2}{2p_2-1}$
11:     Calculate the total charge $c^*(S) \leftarrow n_1^* c_{max} + n_2^* c_{min}$
12:     $\mathcal{L}(S) \leftarrow \mathcal{L}(S) \cup \{f_\ell^*\}$, $\mathcal{C}(S) \leftarrow \mathcal{C}(S) + c^*(S)$
13: return $\mathcal{L}(S), \mathcal{C}(S)$
```

4.3 Privacy-preserving Worker Selection in Platform

After getting the obfuscated information from the requester and worker, the platform needs to perform three steps: information discovery, worker-task matching and worker selection. **Information Discovery:** First, the platform calibrates the obfuscated locations and charges from workers to estimate the number of times workers cover the location ℓ, i.e., $f_\ell^*(S, \epsilon_1)$ in Eq. 3 and the expected total charge $\mathcal{C}(S, \epsilon_2)$ in Eq. 4. In the following, we omit the privacy budget ϵ in notations for simplicity. Then, we propose the Platform Calibration Process (PCP) in Algorithm 4. For a set of workers S, we input the obfuscated location and charge matrices of each worker in S, and get a location matrix $\mathcal{L}(S)$, where each item represents the estimated number of all workers in S cover each location ℓ in $\mathcal{L}$, and the estimated total charge $\mathcal{C}(S)$. Note that $\mathcal{C}(S)$ is a value, not a matrix.

First, we initialize the location matrix $\mathcal{L}(S)$, the charge value $\mathcal{C}(S)$ of S and the probabilities p_1, p_2 (lines 1-2). Then, we count the number of $\ell_i^* = 1$, $c_i^* = c_{max}$ and $c_i^* = c_{min}$ among all obfuscated locations and charges of S (lines 4-6). Then, we calibrate them to estimate their real values based on Eq. 3 (lines 8-10). Note that in line 12, we add f_ℓ^* for each location ℓ to the estimated location matrix $\mathcal{L}(S)$, and $c^*(S)$ indicates the expected charge of S covering ℓ. According to RR, f_ℓ^* is the unbiased estimate of the real value f_ℓ. However, due to the location-charge correlation, the charge confusion is different from the traditional RR. Thus, we prove that the estimated charge $c^*(S)$ is also the unbiased estimate of the real value in Theorem 3.

Theorem 3. *For a specific location ℓ, let $c(S)$ denote the real total charge of workers in S, and $c^*(S)$ denote the estimated total charge by PCP. Then, we have $\mathbb{E}[c^*(S)] = c(S)$, where $\mathbb{E}[c^*(S)]$ is the expected value of $c^*(S)$.*

Proof. As mentioned above, there are two steps that affect the estimation of $c(S)$ for the specific location ℓ.

Location confusion: Due to the location-charge correlation, the location confusion will affect the estimation of $c(S)$. According to the four cases mentioned above, we assume that there are n_1 workers changing from $[1, c_i]$ to $[1, c_i]$, and n_2 workers changing from $[0, 0]$ to $[1, c_i^*]$, where $c_i^* = \mathrm{WCC}((c_{max}+c_{min})/2, \epsilon_2)$ according to WLC. Let $\bar{c}(S)$ denote the mean of all real charges, and $c'(S)$ denote the total charge after the location confusion. Due to the charge discretization in Eq. 7, we have

$$\begin{aligned}
&\mathbb{E}[c'(S)] = \mathbb{E}[\sum\nolimits_{i=1}^{n_1} c_i + \sum\nolimits_{i=1}^{n_2} c_i^*] \\
&= n_1\bar{c}(S) + n_2[c_{max}\cdot(\frac{\bar{c}(S)-c_{min}}{c_{max}-c_{min}}) + c_{min}\cdot(\frac{c_{max}-\bar{c}(S)}{c_{max}-c_{min}})] \\
&= n_1\bar{c}(S) + n_2\cdot\bar{c}(S) = c(S). \qquad (11)
\end{aligned}$$

Note that $\mathbb{E}[\sum_{i=1}^{n_1} c_i] = n_1\cdot\bar{c}(S)$ and $\bar{c}(S) = (c_{max}+c_{min})/2$ due to charges that are drawn from the Gaussian distribution. The last equality holds because the probability of $[1, c_i]$ to $[0, 0]$ is equal to the probability of $[0, 0]$ to $[1, c_i^*]$.

Charge confusion: After the charge discretization in Eq. 7, let n_1', n_2' denote the counts of c_{max}, c_{min} in the real charges. After the charge confusion (lines 2 and 7 in WLC), let $\hat{n}_1, \hat{n}_2$ denote the counts of c_{max}, c_{min} in the obfuscated charges,

$$\mathbb{E}(\hat{n}_1) = n_1'p_2 + n_2'(1-p_2), \mathbb{E}(\hat{n}_2) = n_2'p_2 + n_1'(1-p_2). \qquad (12)$$

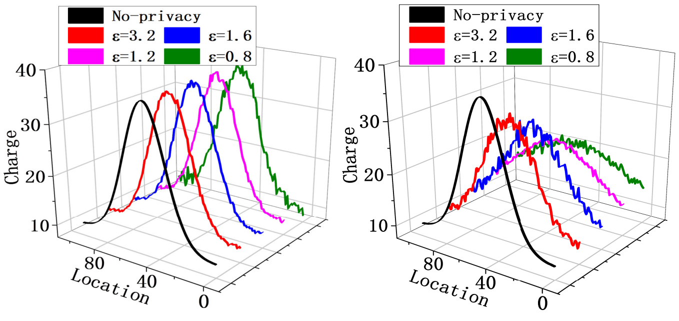
According to PCP (line 10), the counts $\hat{n}_1$ and $\hat{n}_2$ are calibrated based on RR as follows:

$$n_1^* = \frac{(p_2-1)(\hat{n}_1+\hat{n}_2)+\hat{n}_1}{2p_2-1}, n_2^* = \frac{(p_2-1)(\hat{n}_1+\hat{n}_2)+\hat{n}_2}{2p_2-1}. \qquad (13)$$

Based on Eqs. 11-13, we get

$$\begin{aligned}
&\mathbb{E}[c^*(S)] = E[c_{max}n_1^* + c_{min}n_2^*] \\
&= \mathbb{E}[\frac{(c_{max}+c_{min})(p_2-1)(\hat{n}_1+\hat{n}_2) + c_{max}\hat{n}_1 + c_{min}\hat{n}_2}{2p_2-1}] \\
&= [(2p_2-1)c_{max}n_1' + (2p_2-1)c_{min}n_2']/(2p_2-1) \\
&= c_{max}n_1' + c_{min}n_2' = \mathbb{E}[c'(S)] = c(S). \qquad (14)
\end{aligned}$$

Thus, the estimated charge $c^*(S)$ is the unbiased estimate. □



(a) With location-charge correlation (b) Without location-charge correlation

Fig. 4: Location-charge correlation evaluation.

Algorithm 5 Worker-Task Matching (WTM)

```
Input: Estimated worker location matrix $\mathcal{L}(S)$, collection of obfuscated task location matrices $\Gamma$
Output: Utility $F(S, \epsilon_1)$
1:  $F(S, \epsilon_1) \leftarrow 0$
2:  for each $f_\ell^* \in \mathcal{L}(S)$ and $f_\ell^* > 0$ do
3:      $\mathcal{L}^*(S) \leftarrow \{0, \cdots, 0, f_\ell^*, 0, \cdots, 0\}$
4:      for each $\mathcal{L}_j^* \in \Gamma$ do
5:          $\psi \leftarrow \mathcal{L}_j^* \times \mathcal{L}^*(S)$
6:          if $\psi$ is a non-zero symmetric matrix then
7:              $x_\ell \leftarrow 1$
8:          else $x_\ell \leftarrow 0$
9:          $F(S, \epsilon_1) \leftarrow F(S, \epsilon_1) + \min\{f_\ell^*(S, \epsilon_1), 1\} \cdot x_\ell$
10: return $F(S, \epsilon_1)$
```

As we emphasized in Section 4.2, we consider the correlation between location and charge when confusing worker information. Next, we evaluate the benefits of considering the location-charge correlation in Fig. 4. If we obfuscate worker information with location-charge correlation by WLC, the platform calibrates the obfuscated locations and charges from workers in PCP, and the calibration results are shown in Fig. 4a. Specifically, 4a illustrates the estimated charges across locations under different values of the privacy budget. According to [16], we assume that the real charges of workers follow the Gaussian distribution with respect to the locations. Note that other distributions also have similar results. Then, the No-privacy strategy (i.e., the ground truth) shows the Gaussian distribution in Fig. 4a because it has no loss of the correlation due to privacy protection. When the privacy budget ϵ changes, the estimated charges obtained by PCP follow a similar distribution accompanied with slight fluctuations, which means that the platform can recover the information of workers accurately so that workers can be selected reasonably. As a contrast, Fig. 4b shows the calibration results when we obfuscate the worker location and charge without the correlation. We discover that the distribution of estimated charges in Fig. 4b deviates from the real distribution, which means the strategy cannot remain the location-charge correlation and the platform cannot recover the information accurately. Thus, taking the location-charge correlation into account makes the confusion result more reasonable and helps the platform recover the information of workers more accurately as well. This provides support for the following worker-task matching and worker selection.

Worker-task Matching: After the information discovery, the platform has obtained the estimated worker location matrix $\mathcal{L}(S) \leftarrow \mathrm{PCP}(\mathcal{L}_i^*, \mathcal{C}_i^*, \epsilon_1, \epsilon_2)$ of a set of workers S and the collection of obfuscated task location matrices $\Gamma \leftarrow \mathrm{TLC}(\mathcal{L}_j)$. Based on these, the platform conducts the Worker-Task Matching (WTM) process depicted in Algorithm 5 to measure the utility of S. First, we decompose $\mathcal{L}(S)$ into multiple sub-matrices similar to TLC (lines 4-5). Specifically,

Algorithm 6 Worker Selection (WS)

```
Input: Total worker set $U$, budget $B$
Output: Selected worker set $S$
 1: Search all worker sets $S$ with size $|S| \leq 3$ and get a collection $\mathcal{I} = \{S \subseteq U, |S| \leq 3\}$
 2: for each $S' \subseteq \{S|S \subseteq \mathcal{I}, |S| = 3, c(S) \leq B\}$ do
 3:     $t \leftarrow 0, S_t \leftarrow S', U_t \leftarrow U$
 4:     while $U_t \setminus S_t \neq \emptyset$ do
 5:         $t \leftarrow t+1, c_k \leftarrow \mathcal{C}(S_{t-1} \cup \{u_k\}) - \mathcal{C}(S_{t-1})$
 6:         $\delta_t \leftarrow \max_{u_k \in U_{t-1} \setminus S_{t-1}} \frac{F(S_{t-1} \cup \{u_k\}) - F(S_{t-1})}{c_k}$
 7:         $u_t \leftarrow \mathrm{argmax}_{u_k \in U_{t-1} \setminus S_{t-1}} \frac{F(S_{t-1} \cup \{u_k\}) - F(S_{t-1})}{c_k}$
 8:         if $\delta_t \leq 0$ then Terminate the while loop
 9:         if $c(S_{t-1} \cup \{u_t\}) \leq B$ then
10:             $S_t \leftarrow S_{t-1} \cup \{u_i\}, U_t \leftarrow U_{t-1}$
11:         else $S_t \leftarrow S_{t-1}, U_t \leftarrow U_{t-1} \setminus \{u_t\}$
12:     $\mathcal{I} \leftarrow \mathcal{I} \cup S_t$
13: $S \leftarrow \mathrm{argmax}_{S_j \in \mathcal{I}} F(S_j)$
14: return $S$
```

for each item $f^*_\ell \in \mathcal{L}(S)$ and $f^*_\ell > 0$, we construct a new worker location matrix $\mathcal{L}^*(S) \leftarrow \{0, \cdots, 0, f^*_\ell, 0, \cdots, 0\}$, where all items are 0 except the item f^*_ℓ (line 3). Then, for each item in Γ, if there is a validation matrix $\psi = \mathcal{L}^*_j \times \mathcal{L}^*(S)$ that is a non-zero symmetric matrix, it means that workers in S will perform the task in ℓ according to the properties of symmetric matrix. More precisely, if $(\mathcal{L}'_j)^T \times \mathcal{R}^T \times \mathcal{R} \times \mathcal{L}'(S)$ is non-zero symmetric matrix, we obtain $\mathcal{L}'_j = \mathcal{L}'(S)$. Thus, we set $x_\ell = 1$, otherwise $x_\ell = 0$. Next, according to Eq. 5, the platform can measure the utility $F(S, \epsilon_1)$ of S.

Although we have effectively protected the privacy of both the requester and worker, our framework inevitably produces some overhead [36] including three parts: (i) the communication overhead on the requester. As shown in TLC, to obfuscate the locations of tasks, the requester needs to upload n_j+1 matrices to the platform, where n_j denotes the count of ℓ_j=1 in $\mathcal{L}_j$; (ii) the computation overhead on workers. As shown in WLC, each worker will obfuscate his/her locations and charges, whose computation overhead is $O(k^2 \cdot |U|)$; (iii) the computation overhead on the platform. The platform first calibrates the obfuscated information based on PCP, whose computation overhead is $O(k^2)$. Then, the platform conducts the worker-task matching to calculate the utility of a set of workers in WTM with the computation overhead $O(k^2 \cdot n_j)$. Thus, the total computation overhead on the platform is $O(k^2 \cdot n_j)$.

Worker Selection: After the above information discovery and worker-task matching, the platform starts to select workers under the budget constraint. First, we prove that the worker selection problem under the bilateral privacy-preserving framework is NP-hard in Theorem 4 and is non-monotone and non-submodular in Theorem 5.

Theorem 4. *The bilateral privacy-preserving worker selection problem in Eq. 6 is NP-hard.*

Proof. Obviously, when we do not consider the privacy issues of the requester and workers, the problem can be equivalently seen as the problem of maximizing a submodular set function subject to a knapsack constraint [37], which is a well-known NP-hard problem. Hence, the problem in Eq. 6 is NP-hard. □

Theorem 5. *The utility function of a set of workers S is a non-monotone non-submodular function.*

Proof. For simplicity, we only consider the coverage to a location ℓ by a set of workers. Suppose two worker sets $S_1 = \{0, 1\}$ with 2 workers, and $S_2 = \{0, 1, 0\}$ with 3 workers, where 1 means that the worker covers ℓ, and 0 means no coverage. According to the utility function in Eq. 5, let $p_1 = 0.8$, then we have $F(S_1) = 1 \leq F(S_2) = 2/3$. Thus, the utility function is non-monotone. Next, we add a new worker $u_o = 1$ to S_1 and S_2 , and calculate the increment respectively. Then, we get $F(S_1 \cup \{u_o\}) - F(S_1) = 0 < F(S_1 \cup \{u_o\}) - F(S_1) = 1/3$ and $S_1 \subseteq S_2$. According to the submodular property [38], the utility function is non-submodular. □

The worker selection problem proved NP-hard, non-monotone, and non-submodular in Theorems 4-5 is hard to address by existing methods. To this end, we extend the classic non-monotone submodular maximization problem [29] to a non-submodular scenario and design a polynomial-time algorithm referred to as Worker Selection (WS) strategy as in Algorithm 6, which returns an approximate worker set achieving a close performance to the optimal one.

More precisely, given total worker set U and budget B, we search all worker sets with size $|S| \leq 3$ and get an initial collection of worker sets, i.e., $\mathcal{I}$ (line 1). For each set S' in collection $\{S|S \subseteq \mathcal{I}, |S| = 3, c(S) \leq B\}$, we greedily extend it by adding appropriate workers until budget B is exhausted (lines 3-11) and get an extended worker set S_t with size $|S_t| > 3$. Let $S_t = S', U_t = U$. In each step t, we find a worker u_t with the maximum marginal utility δ_t based on the estimated values from PCP. If $\delta_t \leq 0$, the loop part terminates and the current worker set S_t is dropped in order to achieve a lower bound (lines 8). Otherwise, WS adds worker u_t to S_t while satisfying the budget constraint (lines 9-10). Finally, WS returns a worker set S with the maximum utility from collection $\mathcal{I}$ (line 13).

5 Performance Analysis

To analyze the performance of WS, we first introduce Definition 4 [39], [40] to relax the submodularity of the utility function of a set of workers in Eq. 5.

Definition 5. *The γ-submodularity of a non-submodular set function $F()$ is defined as $\gamma = \min_{S_1 \subseteq S_2, u \notin S_2} F_u(S_1)/F_u(S_2)$, where $F_u(S_1) = F(S_1 \cup \{u\}) - F(S_1)$.*

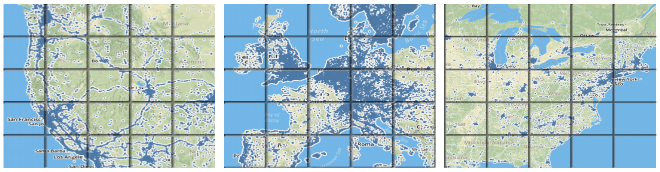
Lemma 1. *Given the γ-submodularity of $F()$, we have $\sum_{u_i \in S_2 \setminus S_1} F_{u_i}(S_1) \geq \gamma \cdot F_{S_2}(S_1)$, where $S_1 \subseteq S_2$.*

Proof. For arbitrary two worker sets S_1 and S_2 with $S_1 \subseteq S_2$ and $S_2 \setminus S_1 = \{u_1, \cdots, u_r\}$, we have

$$
\begin{aligned}
& F_{S_2}(S_1) = F(S_1 \cup S_2) - F(S_1) && (15) \\
& = \sum_{t=1}^{r} [F(S_1 \cup \{u_1, \cdots, u_t\}) - F(S_1 \cup \{u_1, \cdots, u_{t-1}\})] \\
& = \sum_{t=1}^{r} F_{u_t}(S_1 \cup \{u_1, \cdots, u_{t-1}\}) && (16) \\
& \leq (1/\gamma) \cdot \sum_{u_i \in S_2 \setminus S_1} F_{u_i}(S_1),
\end{aligned}
$$

where the last inequality holds due to Definition 4. □

Afterward, based on [38], let S^* denote the optimal worker set of the problem. If $|S^*| \leq 3$, the optimal set will



(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 5: The check-in distribution of three datasets.

TABLE 2: Simulation Parameters

Parameters	Value
Number of grids k	$10 \sim 20$
Privacy budget ϵ_1	$0.1 \sim 0.9$
Privacy budget ϵ_2	0.5
Budget constraint $\mathcal{B}$	$100 \sim 600$
Number of workers $\lvert U\rvert$	$50 \sim 150$
Number of tasks	50
Each worker's total charge	$10 \sim 90$
Charge ratio c_{max}/c_{min}	$1 \sim 6$

be found by WS (line 1). Therefore, we assume that $|S^*| > 3$, and order S^* as $\{u_1, u_2, \cdots, u_t\}$, where

$$u_j = \max_{u\in S^*\setminus\{u_1,u_2,\cdots,u_{j-1}\}} F_{\{u_1,u_2,\cdots,u_{j-1}\}}(\{u\}). \tag{17}$$

The workers are ordered by maximum marginal value without considering their charges. Let $O = \{u_1, u_2, u_3\}$ denote the worker set containing the first three items in S^*. Then, for any worker $u_j \in S^*$, $j \geq 4$ and any worker set $T \subseteq U \setminus \{u_1, u_2, u_3, u_j\}$, according to Definition 4, we have

$$F_{O\cup T}(\{u_j\}) \leq (1/\gamma)\cdot F(\{u_k\}) \leq (1/\gamma)\cdot F(\{u_1\}), \tag{18}$$

$$F_{O\cup T}(\{u_j\}) \leq (1/\gamma)\cdot [F(\{u_1, u_2\}) - F(\{u_1\})], \tag{19}$$

$$F_{O\cup T}(\{u_j\}) \leq (1/\gamma)\cdot [F(\{u_1, u_2, u_3\}) - F(\{u_1, u_2\})]. \tag{20}$$

According to Eqs. 18-20, we can get

$$3F_{O\cup T}(\{u_j\}) \leq (1/\gamma)\cdot F(\{u_1, u_2, u_3\}) = (1/\gamma)\cdot F(O). \tag{21}$$

For the set function $F_O()$, we obtain Lemma 2.

Lemma 2. *For the function $g() = F_O()$, given any worker set $S_1, S_2 \subseteq U$, it holds that*

$$g(S_1\cup S_2) \leq g(S_1) + (1/\gamma)\sum\nolimits_{u_i\in S_1\setminus S_2} [g(S_1\cup\{u_i\}) - g(S_1)]. \tag{22}$$

Proof. According to Lemma 1, we have $g(S_1 \cup S_2) = g(S_1) + g(S_1 \cup S_2) - g(S_1) \leq g(S_1) + 1/\gamma \sum_{u_i\in S_1\setminus S_2} g_{u_i}(S_1) = g(S_1) + 1/\gamma \sum_{u_i\in S_1\setminus S_2}(g(S_1 \cup \{u_i\}) - g(S_1))$. □

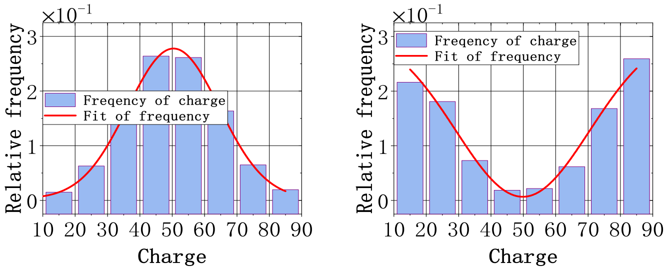
Next, we consider the iteration of WS (lines 2-12), where S_t denotes the selected worker set in the t-iteration in WS, and u_t denotes the selected worker in the t-iteration. Assume that $(t'+1)$-th iteration is the first step in WS when either (i) WS stops (line 8) or (ii) $u_{t'+1} \in S^*$, but $u_{t'+1}$ is dropped by WS. Then, we get Lemma 3.

Lemma 3. $F_O(S_{t'})/F_O(S_{t'} \cup S^*) \geq \gamma(1 - 1/e)$.

Proof. Based on [38] and Lemma 2, we have

$$\frac{F_O(S_{t'})}{F_O(S_{t'}\cup S^*)} \geq \frac{\sum_{j=1}^{B^*}\rho_j}{min_{k\in[1,B^*]}(\sum_{i=1}^{k-1} + (1/\gamma)\cdot B\cdot\rho_k)} \tag{23}$$

$$\geq \gamma\frac{\sum_{j=1}^{B^*}\rho_j}{min_{k\in[1,B^*]}(\sum_{i=1}^{k-1} + B\cdot\rho_k)} \geq \gamma(1 - e^{-B^*/B}) \geq \gamma(1 - e^{-1}),$$



(a) Concave distribution (b) Convex distribution

Fig. 6: Different distributions of worker charges.

where $\gamma \in (0, 1)$, and B, B^* are the temporary variables satisfying $B \leq \mathcal{B} \leq B^*$. □

Lemma 4. *The worker set $S_{t'}$ selected by WS satisfies $F(S_{t'} \cup S^*) \geq F(S^*)$, where S^* is the optimal solution.*

Proof. According to WS (line 8), $F(S_{t'}) \geq 0$. For simplicity, we ignore the min function in Eq. 5, and have $F(S_{t'} \cup S^*) - F(S^*) = \sum_{\ell\in\mathcal{L}}(\frac{(p_1-1)|S_{t'}|}{2p_1-1} + \frac{f(S_{t'})}{2p_1-1}) \geq 0$. The case with the min function can be analyzed in the similar manner. □

Theorem 6. *WS achieves the lower bound $\gamma\cdot(1-1/e)$ with the γ-submodular utility function, where $\gamma \in [0.477, 1)$.*

Proof. According to Definition 4 and Lemmas 1-4, we have

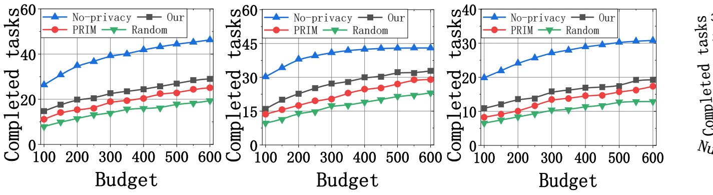
$$\begin{aligned}
F(S_{t'}) &= F(O) + F_O(S_{t'}) \\
&= F(O) + F_O(S_{t'} \cup \{u_{t'+1}\}) - (F(S_{t'} \cup \{u_{t'+1}\}) - F(S_{t'})) \\
&\geq F(O) + \gamma(1 - 1/e)F_O(S_{t'} \cup S^*) - F(O)/(3\gamma) \\
&= \gamma(1 - 1/e)F(S_{t'} \cup S^*) + (1 - 1/(3\gamma) - \gamma(1 - 1/e))F(O) \\
&\geq \gamma(1 - 1/e)F(S_{t'} \cup S^*) \geq \gamma(1 - 1/e)F(S^*),
\end{aligned} \tag{24}$$

where the last inequality holds when $1-1/(3\gamma) \geq \gamma(1-1/e)$, i.e., $\gamma \in [0.477, 1)$. Hence, WS achieves the lower bound $\gamma\cdot(1-1/e)$. In addition, we have proved that the estimates on worker locations and charges are both unbiased estimates in Theorem 3, thus the lower bound is still meaningful. □

6 Simulations

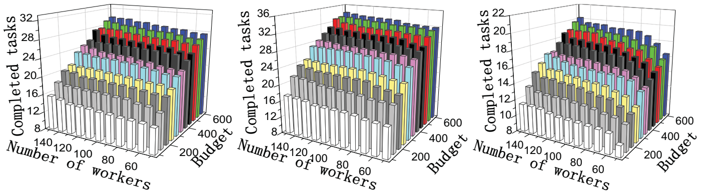
6.1 Trace-based Dataset

In the simulations, we adopt three widely-used real-world check-in datasets: Brightkite [41], Gowalla [41] and Foursquare [42]. The datasets contain a large amount of worker information and PoI information, e.g., the worker id, worker location, venue location, and check-in time. Thus, the datasets are suitable for our simulations, from which we can extract the worker location and task location (PoI) to construct matrices. As shown in Fig. 5, the blue nodes denote the geographic distribution of the workers' locations in three datasets. Specifically, Brightkite includes 58,228 workers and a total of 4,491,143 check-ins from Apr. 2008 to Oct. 2010. Similarly, Gowalla consists of 196,591 workers and 6,442,890 check-ins, and Foursquare contains 456,988 check-ins made by 10,162 workers. In addition, in simulations, we select specific areas based on three datasets and divide each area into $k \times k$ grids according to certain parameters such as task and user density.



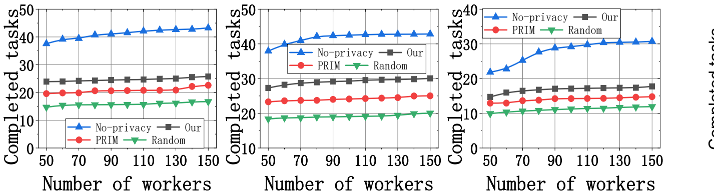
(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 7: Completed tasks vs. Budget.



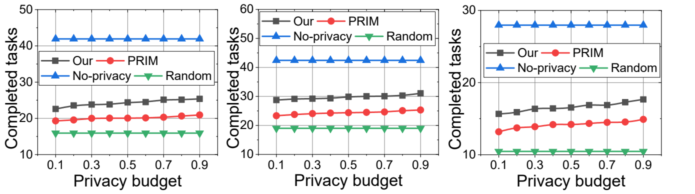
(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 9: Completed tasks vs. Multiple variables.



(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 8: Completed tasks vs. Number of workers.



(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 10: Completed tasks vs. Privacy budget.

6.2 Setup and Metrics

The values of key simulation parameters are described in Table 2. In simulations, we mainly need the worker's location and charge information. For workers' locations, we directly extract each worker's real check-in locations to generate the worker location matrix $\mathcal{L}_i$. For workers' charges, we use a specific distribution to generate each worker's charge for all grids over interval [10, 90]. We select the budget (i.e., $\mathcal{B}$) over the interval [100,600] so as to clearly display the effectiveness of proposed algorithms [43]. Also, we select the privacy budget ϵ_1 over [0.1, 0.9], and $\epsilon_2 = 0.5$ [16]. Based on the number of locations the worker will cover, we get the average charge of going to each location. Using the average charge as the mean of a new distribution, we generate the charge for each location and get each worker's charge matrix $\mathcal{C}_i$. Moreover, we adopt three distributions to generate the worker charge: (1) Uniform distribution; (2) Concave distribution as shown in Fig. 6a, i.e., Gaussian distribution; and (3) Convex distribution as shown in Fig. 6b. Two metrics are used to verify our strategy: completed tasks and privacy leakage.

(1) *Completed tasks*: the number of tasks completed by the selected workers in simulations. Actually, completed tasks refer to the covered tasks since a task is considered to be executed when a mobile worker covers its location.

(2) *Privacy leakage* [13], [44]: the measure of how close the worker's real location matrix is to the obfuscated location matrix following the Kullback-Leibler divergence, and the privacy leakage (PL) in this paper is defined as:

$$PL = 1/(\sum_{u_i \in S} \ln \frac{1}{\Pr(\mathcal{L}_i = \mathcal{L}_i^*)}), \tag{25}$$

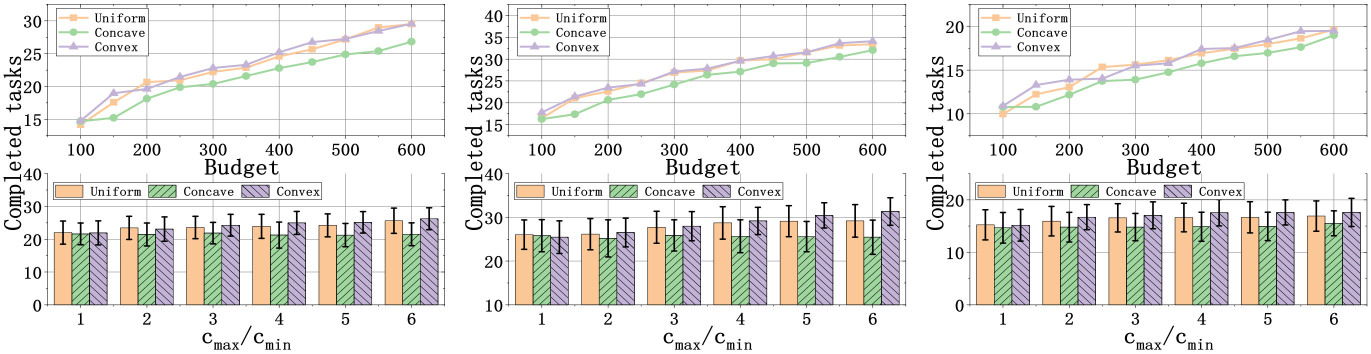
where $\Pr(\mathcal{L}_i = \mathcal{L}_i^*)$ means the probability that the worker's real location matrix $\mathcal{L}_i$ is inferred by the platform. The probability is large when workers are more likely to tell the truth. Note that PL is different from the privacy budget (PB) ϵ in Definition 1. Specifically, PB reflects the privacy level that the mechanism can provide, i.e., how close the initial value is to its obfuscated value. However, PL measures the similarity between the worker's real location matrix and the obfuscated location matrix. Although the definitions of PB and PL are different, they have the following inner-relationship: the higher privacy level (less PB) the mechanism can provide, the less privacy will be leaked to the platform (less PL).

In this paper, we compare our strategy with the following benchmarks: (1) PRIM [13], the state-of-the-art strategy, which exploits the exponentiation mechanism to construct an obfuscation function to protect the worker utility (i.e., the worker's bid) while satisfying the differential privacy. Actually, the worker's bid in PRIM corresponds to the worker's location in this paper, both of which represent the worker's utility. (2) No-privacy, the strategy which is the no-privacy-version of our strategy without the confusion strategies. (3) Random, the strategy which selects workers randomly under the bilateral privacy-preserving framework. (4) No-correlation, the strategy which obfuscates the worker's locations and charges based on RR separately while ignoring the location-charge correlation.

6.3 Evaluation Results

Evaluation of completed tasks: Based on the three datasets, we change the budget from 100 to 600 and conduct simulations in Fig. 7, which shows the results of three datasets: Brightkite, Gowalla, and Foursquare. We observe that the completed tasks of all strategies show an upward trend as the budget grows. This is because, with more payment budget, the platform can select more workers to perform the tasks, then the number of completed tasks grows naturally. We also observe that the No-privacy strategy outperforms the others because No-privacy directly selects workers using the real workers' information and avoids performance loss due to privacy protection. Thus, the selection results are more accurate than using the obfuscated worker information. Moreover, our strategy outperforms PRIM since we use obfuscated information to estimate the real locations and the real charges of a group of workers, instead of directly using obfuscated information to select workers like PRIM. Therefore, our strategy is better than PRIM. In addition, Random strategy obtains the lowest result because of the randomness of the worker selection.

As shown in Fig. 8, we then evaluate the completed tasks with the change of the number of workers from 50 to 150 based on the three datasets. The completed tasks of all strategies grow slightly with the increase of the number



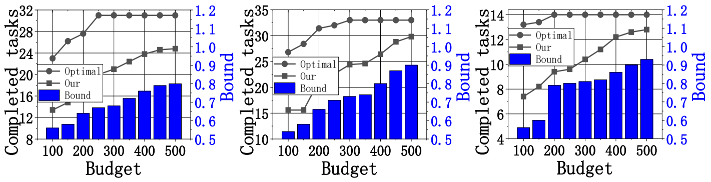
(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 11: Completed tasks vs. Charge distribution & Charge ratio.

of workers because the platform can select suitable workers with high utility and low charge to perform tasks. Similarly, our strategy outperforms others except No-privacy. Then, Fig. 9 demonstrates the overall trend of completed tasks with the changes of both the budget and number of workers.

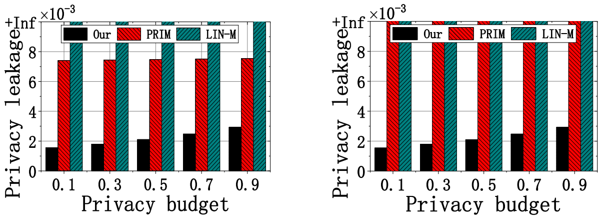
Next, we verify the effect of the privacy budget on completed tasks. As shown in Fig. 10, we investigate the effectiveness of Our, PRIM, No-privacy, and Random as the privacy budget is increased from 0.1 to 0.9. Note that there are two privacy budgets ϵ_1 and ϵ_2 in our strategy (i.e., WLC) for the location confusion and charge confusion, respectively. We focus on the privacy budget ϵ_1 for the location confusion since PRIM only protects workers' utilities (i.e., bid) to satisfy ϵ_1-differential privacy. We observe that as the privacy budget rises, the completed tasks of Our and PRIM increase accordingly. This is because according to Eqs. 1-3, a larger privacy budget indicates a higher probability that the workers tell the truth so that the platform could estimate the statistic of workers more accurately, leading to a better worker selection effect. The larger privacy budget, however, will result in a lower level of privacy protection. Our strategy still has a better performance than PRIM due to the same reasons mentioned above. Note that the results of No-privacy and Random are essentially unchanged as the privacy budget is increased since any noise related to the privacy budget is not considered in these two strategies.

In addition, since worker charge has a great influence on completed tasks, we verify the effect of the charge distribution and charge ratio on completed tasks. First, with the change of the budget, we use three distributions to generate the worker charge and conduct simulations. The results are shown in the upper part of Fig. 11, where Uniform, Concave, and Convex denote the strategies with the corresponding distributions. We discover that all three strategies increase with the growth of the budget. Note that Concave always has the lowest values of completed tasks because concave distribution produces a charge near the mean with a high probability as shown in Fig. 6a. Workers always have a relatively large charge, so the selection effect is not good under the same budget constraint. In the convex distribution as shown in Fig. 6b, the worker charge is either very high or very low, thus it is likely to find the worker with the low charge. Next, we change the value of c_{max}/c_{min} (the mean is not changed) and the results are shown in the lower part of Fig. 11. We can discover that when the difference between c_{max} and c_{min} is small, the values of the three strategies are not much different. But when c_{max}/c_{min} becomes larger, it can be seen that Convex>Uniform>Concave, which is in line with our above statement. The simulations fully explain the effect of charge distribution on worker selection.



(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 12: Completed tasks & Bound vs. Budget.

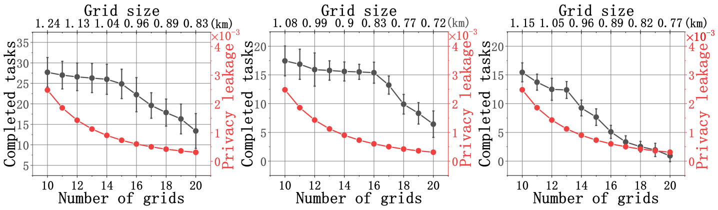


(a) Privacy leakage on locations (b) Privacy leakage on charges

Fig. 13: Privacy leakage vs. Privacy budget.

Evaluation of Lower Bound: Fig. 12 shows the impact of the budget on the lower bound, where Optimal strategy uses the global search to find the optimal worker set under the budget constraint without privacy protection. We discover that the completed tasks of our strategy are quite less than that of Optimal, but as the budget grows, our strategy gets closer to Optimal. This is because when the budget is large enough and the total worker set is not infinite, our strategy will always select enough workers to approach the optimal solution. Moreover, the blue bars denote the values of the lower bound, i.e., the ratio of completed tasks of Our and Optimal. We discover that the values of the lower bound are always greater than 0.5, which is greater than $\gamma_{min}\cdot(1-1/e) = 0.301$. Therefore, the results are consistent with Theorem 6. In addition, the results illustrate that in practical applications, our algorithm is always better than the theoretical lower bound.

Evaluation of Privacy Leakage: According to the definition of the privacy leakage in Eq. 25, we change the privacy budget from 0.1 to 0.9, and the results are shown in Fig. 13, where LIN-M [44] is a privacy-preserving mechanism



(a) Brightkite (b) Gowalla (c) Foursquare

Fig. 14: Completed tasks & Privacy leakage vs. Grid size.

with the trusted platform, that is, all of the worker' information will be exposed to the platform in LIN-M. Fig. 13a demonstrates the privacy leakage while protecting workers' locations, and we discover that the privacy leakage rises with the increase of the privacy budget. As mentioned above, a larger privacy budget means that workers are more likely to tell the truth, which makes the worker selection performance better, but it also leads to more leakage of the worker privacy. Compared with PRIM, our strategy leaks less privacy because we directly protect the worker's location matrix instead of a single value in PRIM. Thus, it is difficult for the platform to guess all the correct locations of a worker. Note that the privacy leakage of LIN-M is positive infinity because workers' real information is exposed completely to the platform, which means that the worker's privacy has been leaked infinitely. Similarly, Fig. 13b shows the trend of the privacy leakage while protecting the worker's charges, where the values of PRIM and LIN-M are both positive infinity because they ignore the privacy protection of the worker's charges. Hence, our strategy achieves the best privacy protection performance.

Evaluation of Number of Grids: In simulations, we select specific areas in cities based on three datasets and divide each area into $k \times k$ grids according to task and user density, where the number of grids is k, and grid size is the real length of the grid (km). Next, we change the number of grids from 10 to 20 and record the completed tasks and privacy leakage as shown in Fig. 14. We discover that with the increase of the number of grids, the completed tasks gradually decline. This is because when we increase the grid number while keeping the task number and worker number unchanged, the locations of workers and tasks become more dispersed so that the previously covered tasks may not be covered now. In addition, with more grids, the accuracy of the platform estimating the worker information in Algorithm 4 has also decreased. More importantly, we find that the privacy leakage also decreases because the worker's location real location matrix is hard to infer by the platform with more grids according to Eq. 25. Actually, as proven in Theorem 1, TLC achieves $\ln k$-Privacy. When the number of grids k increases, the knowledge upper bound brought by the obfuscated information relatively increases, but it does not mean the level of privacy protection is declining. In fact, the level is increasing as shown in Fig. 14. In summary, the privacy leakage is not only related to the privacy budget but also related to the number of girds.

Evaluation of Location-Charge Correlation: In Fig. 4, we have already verified part of the effect of considering the location-charge correlation when confusing the worker information, which can effectively preserve the distribution of charges with respect to locations. Furthermore, we evaluate the impact of considering the location-charge correlation on the accuracy of estimated charges. Note that considering the location-charge correlation or not has no effect on the estimation of workers' locations. We keep other conditions unchanged and change the privacy budget from 0.1 to 0.9, and the results are shown in Table 3, where RE and MSE [16] denote the relative and absolute errors of estimated charges with respect to real charges, respectively. We discover that the values of RE and MSE both decrease with the increase of the privacy budget, which illustrates that as the level of privacy protection decreases, the platform estimates charges more accurately. Moreover, the relative and absolute errors of our strategy are always smaller than No-correlation's, which demonstrates that considering the location-charge correlation while confusing workers' information can effectively improve the platform's estimation accuracy of workers' charges. The reason is that the location-charge correlation is actually the prior knowledge for the charge confusion, and the obfuscated charges could remain more useful information than not considering the correlation.

TABLE 3: Location-charge correlation evaluation.

	ϵ_2	0.1	0.3	0.5	0.7	0.9
Our	RE	0.281	0.215	0.166	0.116	0.075
	log(MSE)	4.303	3.256	2.660	1.930	1.590
No-correlation	RE	0.702	0.590	0.465	0.452	0.448
	log(MSE)	5.722	5.380	5.263	5.259	5.282

7 Conclusion

In this paper, we focused on the privacy concerns in spatial crowdsourcing and proposed a bilateral privacy-preserving framework to select workers under the payment budget constraint without relying on third-party trusted entities. The framework protected the privacy of the requester based on randomized matrix multiplication and also protected the privacy of workers based on randomized response. Specifically, when protecting workers, we proposed a local confusion strategy to jointly protect each worker's locations and charges while utilizing the location-charge correlation, which satisfied local privacy differential. Next, we presented an approximation algorithm to select workers under the bilateral privacy-preserving framework, which achieved a lower bound. Finally, extensive simulations based on real-world datasets verified the performance of our algorithm compared with the state-of-the-art strategy in terms of privacy protection and worker selection.

8 Acknowledge

This work is supported by the National Key R&D Program of China under Grant 2021ZD0112501 and Grant 2021ZD0112502, the National Natural Science Foundations of China under Grant No. 61772230, No. 61972450, No. 62072209, and No. 62102161, the CCF-Baidu Open Fund (No. 2021PP15002000), China Scholarship Council No. 202006170173, and in part by NSF grants CNS 1824440, CNS 1828363, CNS 1757533, CNS 1629746, CNS 1651947, and CNS 1564128.

References

[1] L. Kazemi and C. Shahabi, "Geocrowd: enabling query answering with spatial crowdsourcing," in *Proc. ACM SIGSPATIAL*, 2012.

[2] Y. Tong, L. Chen, and C. Shahabi, "Spatial crowdsourcing: Challenges, techniques, and applications," *PVLDB*, vol. 10, no. 12, pp. 1988–1991, 2017.

[3] Y. Liu, B. Guo, H. Du, Z. Yu, D. Zhang, and C. Chen, "Poster: Foodnet: Optimized on demand take-out food delivery using spatial crowdsourcing," in *Proc. ACM MobiCom*, 2017.

[4] Z. Wang, J. Zhao, J. Hu, T. Zhu, Q. Wang, J. Ren, and C. Li, "Towards personalized task-oriented worker recruitment in mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 2080–2093, 2020.

[5] H. Wang, Y. Yang, E. Wang, W. Liu, Y. XU, and J. Wu, "Combinatorial multi-armed bandit based user recruitment in mobile crowdsensing," in *Proc. IEEE SECON*, 2020.

[6] X. Wang, R. Jia, X. Tian, X. Gan, L. Fu, and X. Wang, "Location-aware crowdsensing: Dynamic task assignment and truth inference," *IEEE Transactions on Mobile Computing*, vol. 19, no. 2, pp. 362–375, 2020.

[7] H. Wang, E. Wang, Y. Yang, J. Wu, and F. Dressler, "Privacy-preserving online task assignment in spatial crowdsourcing: a graph-based approach," in *Proc. IEEE INFOCOM*, 2022.

[8] Z. Wang, X. Pang, J. Hu, W. Liu, Q. Wang, Y. Li, and H. Chen, "When mobile crowdsensing meets privacy," *IEEE Communications Magazine*, vol. 57, no. 9, pp. 72–78, 2019.

[9] Q. Tao, Y. Tong, Z. Zhou, Y. Shi, L. Chen, and K. Xu, "Differentially private online task assignment in spatial crowdsourcing: A tree-based approach," in *Proc. IEEE ICDE*, 2020.

[10] H. To, C. Shahabi, and L. Xiong, "Privacy-preserving online task assignment in spatial crowdsourcing with untrusted server," in *Proc. IEEE ICDE*, 2018.

[11] Z. Wang, J. Hu, R. Lv, J. Wei, Q. Wang, D. Yang, and H. Qi, "Personalized privacy-preserving task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1330–1341, 2018.

[12] 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. ACM WWW*, 2017.

[13] Z. Wang, J. Li, J. Hu, J. Ren, and Y. Li, "Towards privacy-preserving incentive for mobile crowdsensing under an untrusted platform," in *Proc. IEEE INFOCOM*, 2019.

[14] J. Ni, K. Zhang, Q. Xia, X. Lin, and X. S. Shen, "Enabling strong privacy preservation and accurate task allocation for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 19, no. 6, pp. 1317–1331, 2020.

[15] B. Zhao, S. Tang, X. Liu, X. Zhang, and W. Chen, "iTAM: Bilateral privacy-preserving task assignment for mobile crowdsensing," *IEEE Transactions on Mobile Computing*, vol. 20, no. 12, pp. 3351–3366, 2020.

[16] Q. Ye, H. Hu, X. Meng, and H. Zheng, "Privkv: Key-value data collection with local differential privacy," in *Proc. IEEE SSP*, 2019.

[17] S. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[18] G. Yang, Z. Shi, S. He, and J. Zhang, "Socially privacy-preserving data collection for crowdsensing," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 851–861, 2020.

[19] W. Jin, M. Xiao, M. Li, and L. Guo, "If you do not care about it, sell it: Trading location privacy in mobile crowd sensing," in *Proc. IEEE INFOCOM*, 2019.

[20] A. Prorok and V. Kumar, "Privacy-preserving vehicle assignment for mobility-on-demand systems," in *Proc. IEEE IROS*, 2017.

[21] M. Xiao, G. Gao, J. Wu, S. Zhang, and L. Huang, "Privacy-preserving user recruitment protocol for mobile crowdsensing," *IEEE/ACM Transactions on Networking*, vol. 28, no. 2, pp. 519–532, 2020.

[22] X. Ren, C. Yu, W. Yu, S. Yang, X. Yang, J. A. McCann, and S. Y. Philip, "LoPub: high-dimensional crowdsourced data publication with local differential privacy," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 9, pp. 2151–2166, 2018.

[23] T. Wang, X. Yang, X. Ren, W. Yu, and S. Yang, "Locally private high-dimensional crowdsourced data release based on copula functions," *IEEE Transactions on Services Computing*, vol. 15, no. 2, pp. 778–792, 2022.

[24] X. Ren, C. Yu, W. Yu, X. Yang, J. Zhao, and S. Yang, "Dpcrowd: privacy-preserving and communication-efficient decentralized statistical estimation for real-time crowdsourced data," *IEEE Internet of Things Journal*, vol. 8, no. 4, pp. 2775–2791, 2020.

[25] H. Sun, B. Dong, H. Wang, T. Yu, and Z. Qin, "Truth inference on sparse crowdsourcing data with local differential privacy," in *Proc. IEEE International Conference on Big Data*, 2018.

[26] C. Dwork, "Differential privacy," in *Proc. Springer ICALP*, 2006.

[27] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. IEEE FOCS*, 2007.

[28] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: differential privacy for location-based systems," in *Proc. ACM CCS*, 2013.

[29] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE FOCS*, 2013.

[30] Ú. Erlingsson, V. Pihur, and A. Korolova, "RAPPOR: randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM CCS*, 2014.

[31] G. C. Fanti, V. Pihur, and Ú. Erlingsson, "Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries," *Proceedings on Privacy Enhancing Technologies*, vol. 2016, no. 3, pp. 41–61, 2016.

[32] W. Liu, Y. Yang, E. Wang, and J. Wu, "Dynamic user recruitment with truthful pricing for mobile crowdsensing," in *Proc. IEEE INFOCOM*, 2020.

[33] Z. Chen, X. Gao, F. Wu, and G. Chen, "A PTAS to minimize mobile sensor movement for target coverage problem," in *Proc. IEEE INFOCOM*, 2016.

[34] G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, "Privacy-preserving verifiable data aggregation and analysis for cloud-assisted mobile crowdsourcing," in *Proc. IEEE INFOCOM*, 2016.

[35] Z. Wang, X. Pang, Y. Chen, H. Shao, Q. Wang, L. Wu, H. Chen, and H. Qi, "Privacy-preserving crowd-sourced statistical data publishing with an untrusted server," *IEEE Transactions on Mobile Computing*, vol. 18, no. 6, pp. 1356–1367, 2018.

[36] X. Guo, Y. He, X. Zheng, L. Yu, and O. Gnawali, "ZIGFI: harnessing channel state information for cross-technology communication," in *Proc. IEEE INFOCOM*, 2018.

[37] M. Sviridenko, "A note on maximizing a submodular set function subject to a knapsack constraint," *Operations Research Letters*, vol. 32, no. 1, pp. 41–43, 2004.

[38] A. Gupta, A. Roth, G. Schoenebeck, and K. Talwar, "Constrained non-monotone submodular maximization: Offline and secretary algorithms," in *Proc. Springer WINE*, 2010.

[39] C. Qian, J.-C. Shi, Y. Yu, and K. Tang, "On subset selection with general cost constraints," in *Proc. Morgan Kaufmann IJCAI*, 2017.

[40] H. Zhang and Y. Vorobeychik, "Submodular optimization with routing constraints," in *Proc. AAAI*, 2016.

[41] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: user movement in location-based social networks," in *Proc. ACM SIGKDD*, 2011.

[42] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. Magnenat-Thalmann, "Time-aware point-of-interest recommendation," in *Proc. ACM SIGIR*, 2013.

[43] Q. Zhang, Y. Wen, X. Tian, X. Gan, and X. Wang, "Incentivize crowd labeling under budget constraint," in *Proc. IEEE INFOCOM*, 2015.

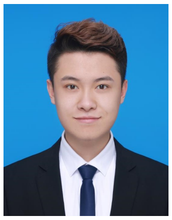
[44] J. Lin, D. Yang, M. Li, J. Xu, and G. Xue, "Frameworks for privacy-preserving mobile crowdsensing incentive mechanisms," *IEEE Transactions on Mobile Computing*, vol. 17, no. 8, pp. 1851–1864, 2018.

Hengzhi Wang received his bachelor's degree in software engineering from Jilin University, Changchun, China, in 2017. Currently, he is studying for the master's degree in computer science and technology from Jilin University, Changchun, China. His current research focuses on mobile crowdsensing and game theory.

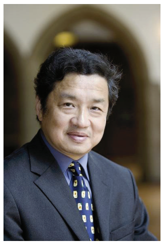
Yongjian Yang received his B.E. degree in automatization from Jilin University of Technology, Changchun, Jilin, China in 1983; his M.E. degree in computer communication from Beijing University of Post and Telecommunications, Beijing, China in 1991; and his Ph.D. in software and theory of computer from Jilin University, Changchun, Jilin, China in 2005. He is currently a professor and a PhD supervisor at Jilin University, the Vice Dean of the Software College of Jilin University, Director of Key lab under the Ministry of Information Industry, Standing Director of the Communication Academy, and a member of the Computer Science Academy of Jilin Province. His research interests include: network intelligence management, wireless mobile communication and services, and wireless mobile communication.

En Wang received his Ph.D. in computer science and technology from Jilin University, Changchun, in 2016. He is currently a Professor in the Department of Computer Science and Technology at Jilin University. He is also a visiting scholar in the Department of Computer and Information Sciences at Temple University in Philadelphia. His current research focuses on the crowdsensin, data mining and mobile computing.

Xiulong Liu is currently a professor in College of Intelligence and Computing, Tianjin University, China. Before that, he received the B.E. and Ph.D. degrees from Dalian University of Technology (China) in 2010 and 2016, respectively. He also worked as a visiting researcher in Aizu University, Japan; a postdoctoral fellow in The Hong Kong Polytechnic University, Hong Kong; and a postdoctoral fellow in the School of Computing Science, Simon Fraser University, Canada. His research interests include wireless sensing and communication, indoor localization, and networking, etc. His research papers were published in many prestigious journals and conferences including TON, TMC, TC, TPDS, TCOM, INFOCOM, and ICNP, etc. He received Best Paper Awards from ICA3PP 2014 and IEEE System Journal 2017. He is also the recipient of CCF Outstanding Doctoral Dissertation award 2017.



Jingxiao Wei received the BE degree in computer science and technology from Jilin University, Changchun, China, in 2019. He is currently pursuing the ME degree in computer science and technology from Jilin University, Changchun, China. He research interests include mobile crowd sensing and privacy protection.



Jie Wu is the Associate Vice Provost for International Affairs at Temple University. He also serves as Director of the Center for Networked Computing and Laura H. Carnell professor in the Department of Computer and Information Sciences. Prior to joining Temple University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Service Computing and the Journal of Parallel and Distributed Computing. Dr. Wu was general cochair/chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, and ACM MobiHoc 2014, as well as program co-chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a CCF Distinguished Speaker and a Fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.