2016 IEEE 36th International Conference on Distributed Computing Systems

Enabling Privacy-Preserving Incentives for Mobile Crowd Sensing Systems

Haiming Jin*, Lu Su†, Bolin Ding‡, Klara Nahrstedt*, Nikita Borisov§
*Department of Computer Science, University of Illinois at Urbana-Champaign, IL, USA
†Department of Computer Science and Engineering, State University of New York at Buffalo, NY, USA
‡Microsoft Research, Redmond, WA, USA
§Department of Electrical and Computer Engineering, University of Illinois at Urbana-Champaign, IL, USA
Email: hjin8@illinois.edu, lusu@buffalo.edu, Bolin.Ding@microsoft.com, {klara, nikita}@illinois.edu

***Abstract*—Recent years have witnessed the proliferation of mobile crowd sensing (MCS) systems that leverage the public crowd equipped with various mobile devices (e.g., smartphones, smartglasses, smartwatches) for large scale sensing tasks. Because of the importance of incentivizing worker participation in such MCS systems, several auction-based incentive mechanisms have been proposed in past literature. However, these mechanisms fail to consider the preservation of workers' *bid privacy*. Therefore, different from prior work, we propose a *differentially private incentive mechanism* that preserves the privacy of each worker's bid against the other honest-but-curious workers. The motivation of this design comes from the concern that a worker's bid usually contains her private information that should not be disclosed. We design our incentive mechanism based on the single-minded reverse combinatorial auction. Specifically, we design a *differentially private*, *approximately truthful*, *individual rational*, and *computationally efficient* mechanism that approximately *minimizes the platform's total payment* with a *guaranteed approximation ratio*. The advantageous properties of the proposed mechanism are justified through not only rigorous theoretical analysis but also extensive simulations.**

***Keywords*-privacy-preserving; incentive mechanism; mobile crowd sensing;**

I. Introduction

The recent proliferation of human-carried mobile devices (e.g., smartphones, smartglasses, smartwatches) with a plethora of on-board sensors (e.g., camera, accelerometer, compass, GPS) has given rise to the emergence of a large variety of people-centric mobile crowd sensing (MCS) systems (e.g., GreenGPS [1], Jigsaw [2], AirCloud [3], and SmartRoad [4]). In a typical MCS system, a central server which is usually a cloud-based platform aggregates and analyzes the sensory data collected by a crowd of diverse participating users, namely (crowd) workers, using their mobile devices. Such MCS systems serve a wide spectrum of applications with significant impact on one's daily live, including healthcare, smart transportation, urban sensing, indoor localization, ambient environment monitoring, etc.

This research was funded in part by the National Science Foundation under award number CNS-1330491, and 1566374. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors. This material is also based upon work supported in part by the Ralph and Catherine Fisher grant.

Participating in such MCS tasks is usually a costly activity for individual workers. The cost depends on various factors including the difficulty of the task, the time a worker spends on executing the sensing tasks, and the amount of system resources (e.g., computing power, battery) that the worker's mobile device consumes. Therefore, without satisfactory rewards that can compensate workers' costs, they will be reluctant to participate in MCS tasks.

Because of the paramount importance of incentivizing worker participation in MCS systems, many reverse auction-based incentive mechanisms [5–16] have been proposed by the research community. In these auctions, a worker submits a bid to the platform containing one or multiple tasks she is interested in and her bidding price for executing these tasks. Based on workers' bids, the platform acting as the auctioneer determines the winners who are assigned to execute the tasks they bid and the payments paid to the selected winners. Furthermore, designing a truthful auction where every worker bids to the platform her true interested tasks and the corresponding true task execution cost is a common objective.

However, all the aforementioned incentive mechanisms [5–16] fail to consider the preservation of workers' *bid privacy*. Although the platform is usually considered to be trusted, there exist some *honest-but-curious* workers who strictly follow the protocol of the MCS system, but try to infer information about other workers' bids. A worker's bid usually contains her private and sensitive information. For example, a worker's bidding task set could imply her personal interests, knowledge base, etc. In geotagging MCS systems that provide accurate localization of physical objects (e.g., automated external defibrillator [17], pothole [18]), bidding task sets contain the places a worker has visited or will visit, the disclosure of which breaches her location privacy. Similar to bidding task set, a worker's bidding price could also be utilized to infer her sensitive information. For example, bidding price could imply the type of mobile devices a worker uses for an MCS task, because usually workers tend to bid more if their mobile devices are more expensive.

Typically, the change in one worker's bid has the potential to shift the overall payment profile (i.e., payments to all

1063-6927/16 $31.00 © 2016 IEEE
DOI 10.1109/ICDCS.2016.50

IEEE computer society

workers) significantly. It is possible that a curious worker could infer information about other workers' bids from the different payments she receives in two rounds of the auction. To address this issue, we incorporate the notion of *differential privacy* [19–22], which ensures that the change in any worker's bid will not bring a significant change to the resulting payment profile. Therefore, different from all existing incentive mechanisms for MCS systems, we design a *differentially private incentive mechanism that protects workers' bid privacy against honest-but-curious workers*.

Because of workers' *selfish* and *strategic* behaviours that aim to maximize their own utilities and the combinatorial nature of the tasks executed by each worker, we design an incentive mechanism based on the *single-minded reverse combinatorial auction*. In our mechanism, every worker bids on a set of tasks that she is interested to execute. The platform serves as the auctioneer and determines the winners and the payment profile that *minimize its total payment to all the winners*. In sum, this paper has the following contributions.

- Different from all existing incentive mechanisms for MCS systems, we design a *differentially private* incentive mechanism that preserves the privacy of each worker's bid against the other honest-but-curious workers.
- Apart from differential privacy, our mechanism also satisfies the desirable economic properties of approximate truthfulness and individual rationality.
- Algorithmically, our mechanism is computationally efficient and minimizes the platform's total payment with a guaranteed approximation ratio.

II. Related Work

Game theoretic models [5–16, 23–25] have been widely utilized in designing incentive mechanisms for MCS systems because of their ability to capture and tackle workers' strategic behaviors. Among them, one major category is auction-based incentive mechanisms [5–16].

Yang *et al.* [8] propose an auction-based user-centric incentive mechanism, which does not consider workers' misreporting of bidding task sets. Zhang *et al.* [6] design an incentive mechanism tailored for crowd labeling tasks under the platform's budget constraint. Zhang *et al.* [7] incorporate both the cooperation and competition among participating workers. Feng *et al.* [9] aim to minimize the social cost in their mechanism. Furthermore, [10, 11] design quality of information aware incentive mechanisms, [12, 13] design incentive mechanisms where workers' task execution costs are known prior information to the platform, [5] studies providing long-term participating incentive to crowd workers and [14–16] design online incentive mechanisms for MCS systems where workers arrive sequentially.

However, all the aforementioned existing work fail to consider the preservation of workers' privacy. In contrast, we incorporate the notion of *differential privacy* [19–22] and design a differentially private incentive mechanism for MCS systems that protects *workers' bid privacy*. There do exist several related work [26–29] regarding privacy-preserving incentive mechanisms for MCS systems. Instead of bid privacy, [29] focuses on protecting workers' privacy leakage from the aggregated data. [26–28] do not consider workers' strategic behaviours, and do not use auction-based incentive mechanisms. Instead, they adopt credit systems [26, 27] and untraceable electronic currency [28].

Another line of related work [20–22, 30] designs privacy-preserving auctions for various different applications. Encrypting workers' bids in [30] does not resolve the issue of curious workers' inferring information about other workers' bids from the payments they receive. The differentially private auction frameworks [20–22] designed for forward auctions cannot be directly applied in the reverse auction scenario considered in this paper.

III. Preliminaries

In this section, we present an overview of MCS systems, the aggregation method, our auction model, and design objectives.

A. System Overview

The MCS system considered in this paper consists of a cloud-based platform and a set of N participating workers denoted as $\mathcal{N} = \{w_1, \cdots, w_N\}$.

In this paper, we are particularly interested in MCS systems that host a set of K classification tasks, denoted as $\mathcal{T} = \{\tau_1, \cdots, \tau_K\}$, namely ones that require workers to locally decide the classes of the objects or events she has observed, and report her local decisions (i.e., labels of the observed objects or events) to the platform. Here, we assume that all tasks in $\mathcal{T}$ are binary classification tasks, which constitute a significant portion of the tasks posted on MCS platforms. Examples of such tasks include tagging whether or not a segment of road surface has potholes or bumps [18, 31], labeling whether or not traffic congestion happens at a specific road segment [32], etc. Each binary classification task $\tau_j \in \mathcal{T}$ has a true class label l_j, unknown to the platform, which is either $+1$ or -1. If worker w_i is selected to execute task τ_j, she will provide a label $l_{i,j}$ to the platform.

Currently, a major challenge in designing reliable MCS systems lies in the fact that the sensory data provided by individual workers are usually unreliable due to various reasons including carelessness, background noise, lack of sensor calibration, poor sensor quality, etc. To overcome this issue, the platform has to aggregate the labels provided by multiple workers, as this will likely cancel out the errors of individual workers and infer the true label. We describe the workflow of the MCS system as follows.

- The platform firstly announces the set of binary classification tasks, $\mathcal{T}$, to the workers.
- Then, the workers and the platform start the auctioning stage, where the platform acts as the *auctioneer* purchasing the labels provided by the workers. Every worker w_i submits her bid $b_i = (\Gamma_i, \rho_i)$, which is a tuple consisting of the set of tasks Γ_i she wants to execute and her bidding price ρ_i for providing labels about these tasks. We use $\mathbf{b} = (b_1, \cdots, b_N)$ to denote workers' bid profile.
- Based on workers' bids, the platform determines the set of winners (denoted as $\mathcal{S} \subseteq \mathcal{N}$) and the payment p_i paid to each worker w_i. We use $\mathbf{p} = (p_1, \cdots, p_N)$ to denote workers' payment profile.
- After the platform aggregates workers' labels to infer the true label of every task, it gives the payment to the corresponding winners.

Every worker w_i has a *skill level* $\theta_{i,j} \in [0, 1]$ for task τ_j, which is the probability that the label $l_{i,j}$ provided by worker w_i about task τ_j equals to the true label l_j, i.e., $\Pr[l_{i,j} = l_j] = \theta_{i,j}$. We use the matrix $\Theta = [\theta_{i,j}] \in [0, 1]^{N \times K}$ to denote the skill level matrix of all workers. We assume that the platform maintains a historical record of the skill level matrix Θ utilized as one of the inputs for winner and payment determination. There are many methods that the platform could use to estimate Θ. In the cases where the platform has access to the true labels of some tasks a *priori*, it can assign these tasks to workers in order to estimate Θ as in [33]. When ground truth labels are not available, Θ can still be effectively estimated from workers' previously submitted data using algorithms such as those in [34–38]. Alternatively, in many applications Θ can be inferred from some explicit characteristics of the workers (e.g., a worker's reputation and experience of executing certain types of sensing tasks, the type and price of a worker's sensors) using the methods proposed in [39]. The issue of exactly which method is used by the platform to calculate Θ is application dependent and out of the scope of this paper.

B. Aggregation Method

In this paper, we reasonably assume that the platform uses a weighted aggregation method to calculate the aggregated label $\widehat{l}_j$ for each task τ_j based on the collected labels. That is, $\widehat{l}_j = \text{sign}(\sum_{i:w_i \in \mathcal{S}, \tau_j \in \Gamma_i} \alpha_{i,j} l_{i,j})$, where $\alpha_{i,j}$ is the weight corresponding to the label $l_{i,j}$. In fact, many sophisticated state-of-the-art data aggregation mechanisms, such as those proposed in [34–38], also adopt the weighted aggregation method to calculate the aggregation results. Given the aggregation method, the platform selects winners so that the aggregation error of each task τ_j's label is upper bounded by a predefined threshold δ_j. That is, the platform aims to ensure that $\Pr[\widehat{l}_j \neq l_j] \leq \delta_j$ holds for every task $\tau_j \in \mathcal{T}$. We directly apply in this paper the results derived in [40], formally summarized in Lemma 1, regarding the relationship between the selected winners' skill levels and the upper bounds of tasks' aggregation error.

Lemma 1. *Suppose the platform utilizes a weighted aggregation algorithm that calculates the aggregated label $\widehat{l}_j$ of task $\tau_j \in \mathcal{T}$ according to $\widehat{l}_j = \text{sign}(\sum_{i:w_i \in \mathcal{S}, \tau_j \in \Gamma_i} \alpha_{i,j} l_{i,j})$. Thus, $\Pr[\widehat{l}_j \neq l_j] \leq \delta_j$ holds if and only if $\alpha_{i,j} = 2\theta_{i,j} - 1$ and*

$$\sum_{i:w_i \in \mathcal{S}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2 \geq 2 \ln \Big(\frac{1}{\delta_j}\Big), \tag{1}$$

where $\delta_j \in (0, 1)$.

We refer to Equation 1 as the *error bound constraint* in the rest of this paper. Essentially, Lemma 1 presents a necessary and sufficient condition for $\Pr[\widehat{L}_j \neq l_j] \leq \delta_j$ to hold ($\forall \tau_j \in \mathcal{T}$) for a weighted aggregation algorithm. That is, the aggregated label $\widehat{l}_j$ should be calculated as $\widehat{l}_j = \text{sign}\big(\sum_{i:w_i \in \mathcal{S}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1) l_{i,j}\big)$ and the sum of the value $(2\theta_{i,j} - 1)^2$'s for all winner w_i's that execute task τ_j should not be smaller than the threshold $2 \ln \big(\frac{1}{\delta_j}\big)$. Intuitively, the larger the value $(2\theta_{i,j} - 1)^2$ is, the more informative the label $l_{i,j}$ will be to the platform. When the value $(2\theta_{i,j} - 1)^2$ approaches 0, or equivalently $\theta_{i,j}$ approaches 0.5, the label $l_{i,j}$ will be closer to a random noise.

C. Auction Model

In the rest of the paper, we will refer to any subset of tasks of $\mathcal{T}$ as a *bundle*. Since in the MCS system considered in this paper every worker bids on one bundle of tasks, we use *single-minded reverse combinatorial auction with heterogeneous cost (hSRC auction)*, formally defined in Definition 1, to model the problem.

Definition 1 (hSRC Auction)**.** *We define the single-minded reverse combinatorial auction with heterogeneous cost, namely hSRC auction, as follows. In the hSRC auction, any worker w_i has a set of K_i possible bidding bundles denoted as $\mathcal{T}_i = \{\Gamma_{i,1}, \cdots, \Gamma_{i,K_i}\}$. For providing labels about all the tasks in each bundle $\Gamma_{i,k} \in \mathcal{T}_i$, the worker has a cost $c_{i,k}$. Furthermore, every worker w_i is only interested in one of the bundles in $\mathcal{T}_i$, denoted as Γ_i^* with cost c_i^*.*

Noted that the hSRC auction defined in Definition 1 is a generalization of traditional single-minded combinatorial auctions, such as those in [10, 41, 42]. Typically, in traditional single-minded combinatorial auctions, all the possible bidding bundles of a worker have the same cost. However, in our hSRC auction, the cost $c_{i,k}$'s for every bundle $\Gamma_{i,k} \in \mathcal{T}_i$ do not necessarily have to be the same. In MCS systems, workers usually have different costs for executing different bundles, which makes our definition of hSRC auction more suitable to the problem studied in this paper. In Definition 2, we define a worker's truthful bid.

Definition 2 (Truthful Bid)**.** *We define bid $b_i^* = (\Gamma_i^*, c_i^*)$ which contains worker w_i's true interested bundle Γ_i^* and the corresponding cost c_i^* as her truthful bid.*

In Definition 3 and 4, we present the formal definitions of a worker's utility and the platform's total payment.

Definition 3 (Worker's Utility)**.** *Suppose a worker w_i bids $\Gamma_{i,k} \in \mathcal{T}_i$ in the hSRC auction. If she is a winner, she will be paid p_i by the platform. Otherwise, she will not be allocated any task and receives zero payment. Therefore, the utility of the worker w_i is*

$$u_i = \begin{cases} p_i - c_{i,k}, & \text{if } w_i \in \mathcal{S} \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

Definition 4 (Platform's Payment)**.** *The platform's total payment to all workers given the payment profile $\mathbf{p}$ and the winner set $\mathcal{S}$ is*

$$R(\mathbf{p}, \mathcal{S}) = \sum_{i:w_i \in \mathcal{S}} p_i. \tag{3}$$

D. Design Objective

Since workers are strategic in our hSRC auction, it is possible that a worker could submit a bid different from the truthful bid defined in Definition 2 in order to obtain more utility. To address this problem, one of our goals is to design a *truthful* mechanism, where every worker maximizes her utility by bidding her truthful bid regardless of other workers' bids. In practice, ensuring exact truthfulness for the hSRC auction is too restrictive. Therefore, we turn to a weaker but more practical notion of *γ-truthfulness in expectation* [20, 43], formally defined in Definition 5.

Definition 5 (γ-truthfulness)**.** *An hSRC auction is γ-truthful in expectation, or γ-truthful for short, if and only if for any bid $b_i \neq b_i^*$ and any bid profile of other workers $\mathbf{b}_{-i}$, there is*

$$\mathbb{E}\big[u_i(b_i^*, \mathbf{b}_{-i})\big] \geq \mathbb{E}\big[u_i(b_i, \mathbf{b}_{-i})\big] - \gamma, \tag{4}$$

where γ is a small positive constant.

γ-truthfulness ensures that no worker is able to make more than a slight γ gain in her expected utility by bidding untruthfully. Therefore, we reasonably assume that each worker w_i would bid her truthful bid b_i^*, if our hSRC auction satisfies γ-truthfulness. Apart from γ-truthfulness, another desirable property of our hSRC auction is *individual rationality*, which implies that no worker has negative utility. This property is crucial in that it prevents workers from being disincentivized by receiving negative utilities. We formally define this property in the following Definition 6.

Definition 6 (Individual Rationality)**.** *An hSRC auction is individual rational if and only if $u_i \geq 0$ holds for every worker $w_i \in \mathcal{N}$.*

Simply paying workers according to the output payment profile of the auction poses threats to the privacy of workers' bids. Because the change in one worker's bid has the potential to shift the payment profile significantly, it is possible for a curious worker to infer other workers' bids from the different payments she receives in two rounds of auction. Therefore, we aim to design a differentially private mechanism [19–22], formally defined in Definition 7.

Definition 7 (Differential Privacy)**.** *We denote the proposed hSRC auction as a function $M(\cdot)$ that maps an input bid profile $\mathbf{b}$ to a payment profile $\mathbf{p}$. Then, $M(\cdot)$ is ϵ-differentially private if and only if for any possible set of payment profiles $\mathcal{A}$ and any two bid profiles $\mathbf{b}$ and $\mathbf{b}'$ that differ in only one bid, we have*

$$\Pr\big[M(\mathbf{b}) \in \mathcal{A}\big] \leq \exp(\epsilon)\Pr\big[M(\mathbf{b}') \in \mathcal{A}\big], \tag{5}$$

where ϵ is a small positive constant usually referred to as privacy budget.

Differential privacy ensures that the change in any worker's bid will not bring a significant change to the resulting payment profile. Hence, it is difficult for the curious workers to infer information about other workers' bids from the outcome (i.e., payment profile) of the mechanism. In this paper, to achieve differential privacy we introduce randomization to the outcome of our mechanism, similar to [20–22].

In short, we aim to design a *γ-truthful, individual rational and ϵ-differentially private* incentive mechanism in this paper.

IV. Mathematical Formulation

In this section, we present our formal mathematical problem formulation.

In this paper, we adopt the natural and commonly used *optimal single-price payment*, as in [21, 44, 45], as our optimal payment benchmark, because it is within a constant factor of the payment of any mechanism with price differentiation, as proved in [45]. Therefore, we aim to design a single-price mechanism that pays every winner in $\mathcal{S}$ according to the same price p.

To simplify our analysis, we assume that the possible values of the cost $c_{i,k}$ for a worker w_i to execute a bundle of tasks $\Gamma_{i,k} \in \mathcal{T}_i$ forms a *finite set* $\mathcal{C}$. The smallest and largest element in $\mathcal{C}$ is $c_{\min}$ and $c_{\max}$ respectively. Given the winner set $\mathcal{S}$, for an individual rational single-price mechanism, the platform's total payment is minimized if and only if the price p equals to the largest cost of the workers in $\mathcal{S}$, that is $p = \max_{w_i \in \mathcal{S}} c_{i,k}$. This is because otherwise the platform can always let $p = \max_{w_i \in \mathcal{S}} c_{i,k}$ and obtains a smaller total payment while maintaining individual rationality. Therefore, the set $\mathcal{P}$ containing all possible prices should satisfy that $\mathcal{P} \subseteq \mathcal{C}$. Furthermore, we define that a price p is feasible if and only if it is possible to select a set of winners $\mathcal{S}$ among the workers with bidding prices $\rho_i \leq p$ such that the error

bound constraint defined in Equation 1 is satisfied for every task. Then, we define the price set $\mathcal{P}$ as the set containing all values in the set $\mathcal{C}$ that are feasible. Thus, obviously we have $c_{\max} \in \mathcal{P} \subseteq \mathcal{C}$. Given a price p and all the other parameters, we use $S_{\text{OPT}}(\cdot)$ to denote the mechanism that maps p to the minimum-cardinality winner set such that every task's error bound constraint is satisfied. Thus, the optimal total payment R_{OPT} can be written as

$$R_{\text{OPT}} = \min_{p \in \mathcal{P}} p|S_{\text{OPT}}(p)|. \tag{6}$$

Therefore, given a price p, the total payment minimization (TPM) problem can be formulated as the following integer linear program.

TPM Problem:

$$\min \sum_{i:w_i \in \mathcal{N}'} p x_i \tag{7}$$

$$\text{s.t.} \sum_{i:w_i \in \mathcal{N}', \tau_j \in \Gamma_i} q_{i,j} x_i \geq Q_j, \qquad \forall \tau_j \in \mathcal{T} \tag{8}$$

$$x_i \in \{0,1\}, \qquad \forall w_i \in \mathcal{N}' \tag{9}$$

Constants. The TPM problem takes as inputs a given price p, workers' bid profile $\mathbf{b}$, the matrix $\mathbf{q}$, the vector $\mathbf{Q}$, the task set $\mathcal{T}$ and the set $\mathcal{N}' = \{w_i | w_i \in \mathcal{N}, \rho_i \leq p\}$ with cardinality N' containing all the workers whose bidding prices are not larger than p.

Variables. In the TPM problem, we have a vector of N' binary variables $\mathbf{x} = (x_1, \cdots, x_{N'})$. For every worker $w_i \in \mathcal{N}'$, there is a binary variable x_i indicating whether this worker is in the winner set $\mathcal{S}$. That is, $x_i = 1$ if $w_i \in \mathcal{S}$ and $x_i = 0$ if $w_i \notin \mathcal{S}$.

Objective function. Based on the definition of variables $\mathbf{x}$, $\sum_{i:w_i \in \mathcal{N}'} x_i$ equals to the cardinality of the winner set $\mathcal{S}$. Therefore, given a price p, the objective function $\sum_{i:w_i \in \mathcal{N}'} p x_i$ represents the platform's total payment to all the winners.

Constraints. For simplification of presentation, we introduce the following notations. $q_{i,j} = (2\theta_{i,j} - 1)^2$, $Q_j = 2\ln\left(\frac{1}{\delta_j}\right)$, $\mathbf{q} = [q_{i,j}] \in [0,1]^{N \times K}$ and $\mathbf{Q} = (Q_1, \cdots, Q_K)$. Thus, Constraint 8 is equivalent to the error bound constraint represented by Equation 1 in Lemma 1, which ensures that the aggregation error of every task $\tau_j \in \mathcal{T}$ is not larger than a threshold δ_j.

In Theorem 1, we prove the NP-hardness of the TPM problem.

Theorem 1. *The TPM problem is NP-hard.*

Proof: Since p is a constant, the TPM problem has the same computational complexity as the modified TPM problem that minimizes $\sum_{i:w_i \in \mathcal{N}'} x_i$ with the same set of constraints. Thus, we turn to prove the NP-hardness of the modified TPM problem, instead.

We start our proof by introducing an instance of the minimum set cover (MSC) problem with a universe of K elements $\mathcal{U} = \{\tau_1, \cdots, \tau_K\}$ and a set of N sets $\mathcal{H} = \{\Gamma_1, \cdots, \Gamma_N\}$. The objective of the MSC problem is to find the minimum-cardinality subset of $\mathcal{H}$ whose union contains all the elements in $\mathcal{U}$. We construct an instance of the modified TPM problem based on this instance of the MSC problem. Firstly, we construct Γ'_i from Γ_i where every $\tau_j \in \Gamma_i$ has $h_{i,j} \in \mathbb{Z}^+$ copies in Γ'_i. Furthermore, we require that the selected sets cover every $\tau_j \in \mathcal{U}$ for at least H_j times. Therefore, we get an instance of the modified TPM problem where $\mathbf{q} = [h_{i,j}] \in (\mathbb{Z}^+)^{N \times K}$, $\mathbf{Q} = (H_1, \cdots, H_K)$ and the bidding bundle profile $\mathbf{\Gamma} = (\Gamma'_1, \cdots, \Gamma'_N)$. In fact, the modified TPM problem represents a richer family of problems where elements in $\mathbf{q}$ and $\mathbf{Q}$ can be positive real values. Therefore, every instance of the NP-complete MSC problem is polynomial-time reducible to the modified TPM problem. The modified TPM problem, and equivalently the TPM problem, is NP-hard. ■

V. Mechanism Design

Because of the NP-hardness of the TPM problem shown in Theorem 1, even given the price p, it is impossible to calculate in polynomial time the set of winners that minimize the platform's total payment unless P = NP. Let alone we eventually need to select an optimal price from the price set $\mathcal{P}$. Therefore, we aim to design a *polynomial-time* mechanism that gives us *an approximately optimal total payment* with *a guaranteed approximation ratio* to the optimal total payment R_{OPT}. In addition, we also take into consideration the bid privacy preserving objective when designing the mechanism. We present our mechanism in Algorithm 1, namely *differentially private hSRC (DP-hSRC) auction*, that satisfies all our design objectives.

Algorithm 1 takes as inputs the privacy budget ϵ, the cost upper bound $c_{\max}$, the worker set $\mathcal{N}$, the task set $\mathcal{T}$, the price set $\mathcal{P}$, workers' bid profile $\mathbf{b}$, the $\mathbf{q}$ matrix and the $\mathbf{Q}$ vector. It outputs the winner set $\mathcal{S}$ and the payment p paid to each winner. Firstly, it sorts workers according to the ascending order of their bidding prices such that $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_N$ (line 1). Then, it initializes several parameters (line 2-5). It finds the minimum price $p_{\min}$ in $\mathcal{P}$ (line 2) and the index $i_{\min}$ of the largest bidding price that does not exceed $p_{\min}$ (line 3). The algorithm constructs an index set $\mathcal{I}$ containing all the integers from $i_{\min}$ to N (line 4). Set $\mathcal{I}$ contains every worker index i such that a winner set $\mathcal{S}_i$ is calculated among the workers with bidding prices that are not larger than ρ_i. In the last step of the initialization, the algorithm creates an extra bidding price ρ_{N+1} by adding a small positive constant δ to $c_{\max}$ (line 5) to ensure that ρ_{N+1} is greater than $\forall p \in \mathcal{P}$. The purpose of creating ρ_{N+1} is to make sure that every price $p \in \mathcal{P}$ is considered by line 14 and 15 in the main loop (line 6-15) for exactly once.

After the initialization phase, Algorithm 1 calculates the winner set for every possible price $p \in \mathcal{P}$ (line 6-15). Intuitively, we need to calculate the winner set for every given

Algorithm 1: DP-hSRC Auction

Input: ϵ, $c_{\max}$, $\mathbf{b}$, $\mathbf{q}$, $\mathbf{Q}$, $\mathcal{N}$, $\mathcal{T}$, $\mathcal{P}$;
Output: $\mathcal{S}$, p;

1 sort workers according to the ascending order of bidding prices such that $\rho_1 \leq \rho_2 \leq \cdots \leq \rho_N$;
// Initialization
2 $p_{\min} \leftarrow \min_{p \in \mathcal{P}} p$;
3 $i_{\min} \leftarrow \arg\max_{i:\rho_i \leq p_{\min}} \rho_i$;
4 $\mathcal{I} \leftarrow \{i_{\min}, i_{\min}+1, \cdots, N\}$;
// Add a small constant $\delta > 0$ to $c_{\max}$
5 $\rho_{N+1} \leftarrow c_{\max} + \delta$;
// Calculates the winner sets
6 **foreach** $i \in \mathcal{I}$ **do**
7 $\quad \mathcal{S}_i \leftarrow \emptyset$, $\mathbf{Q}' \leftarrow \mathbf{Q}$, $\mathcal{N}' \leftarrow \{w_k | \rho_k \leq \rho_i\}$;
8 $\quad$ **while** $\sum_{j:\tau_j \in \mathcal{T}} Q'_j \neq 0$ **do**
9 $\quad\quad i_{\max} = \arg\max_{i:w_i \in \mathcal{N}'} \sum_{j:\tau_j \in \Gamma_i} \min\{Q'_j, q_{i,j}\}$;
10 $\quad\quad \mathcal{S}_i \leftarrow \mathcal{S}_i \cup \{w_{i_{\max}}\}$;
11 $\quad\quad \mathcal{N}' \leftarrow \mathcal{N}' \setminus \{w_{i_{\max}}\}$;
$\quad\quad$ // Update the residual $\mathbf{Q}'$ vector
12 $\quad\quad$ **foreach** j *s.t.* $\tau_j \in \mathcal{T}$ **do**
13 $\quad\quad\quad Q'_j \leftarrow Q'_j - \min\{Q'_j, q_{i_{\max},j}\}$;
$\quad$ // Assign the same winner set $\mathcal{S}_i$ to every possible price in $[\rho_i, \rho_{i+1})$
14 $\quad$ **foreach** $p \in \mathcal{P} \cap [\rho_i, \rho_{i+1})$ **do**
15 $\quad\quad S(p) \leftarrow \mathcal{S}_i$;
16 randomly pick a price p according to the distribution

$$\Pr[p = x] = \frac{\exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big)}{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S(y)|}{2N c_{\max}}\Big)},\ \forall x \in \mathcal{P};$$

// Obtain the corresponding winner set
17 $\mathcal{S} \leftarrow S(p)$;
18 **return** $\{\mathcal{S}, p\}$;

price $p \in \mathcal{P}$. However, for all possible prices between two consecutive bidding prices, that is $\forall p \in \mathcal{P} \cap [\rho_i, \rho_{i+1})$, the winner sets are the same. Therefore, to reduce the computational complexity and remove its dependency on the number of possible prices (i.e., $|\mathcal{P}|$), we only need to calculate the winner set for every price $p \in \{\rho_{i_{\min}}, \rho_{i_{\min}+1}, \cdots, \rho_N\}$. At the beginning of every iteration of the main loop (line 6-15), Algorithm 1 initializes the winner set $\mathcal{S}_i$ as $\emptyset$, the residual $\mathbf{Q}'$ vector as $\mathbf{Q}$ and the candidate winner set $\mathcal{N}'$ as every worker w_k with bidding price ρ_k that is not larger than ρ_i (line 7). The inner loop (line 8-13) is executed until the error bound constraints for all tasks are satisfied, or equivalently until $\mathbf{Q}' = \mathbf{0}^{K \times 1}$. In every iteration of the inner loop (line 8-13), the worker $w_{i_{\max}}$ that provides the most improvement to the feasibility of Constraint 8 is selected as the new winner (line 9). Hence, $w_{i_{\max}}$ is included in $\mathcal{S}_i$ (line 10) and excluded from $\mathcal{N}'$ (line 11). After $w_{i_{\max}}$ is selected, the algorithm updates the residual $\mathbf{Q}'$ vector (line 12-13).

To ensure differential privacy, we introduce randomization to the output price. We extend the exponential mechanism proposed in [20] and set the probability that the output price p of Algorithm 1 equals to a price $x \in \mathcal{P}$ to be proportional to the value $\exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big)$. That is,

$$\Pr[p = x] \propto \exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big),\ \forall x \in \mathcal{P}. \quad (10)$$

One important rationale of setting the probability of every possible price as the form in Equation 10 is that the price resulting in a smaller total payment will have a larger probability to be sampled. In fact, the probability increases exponentially with the decrease of the total payment and the distribution is substantially biased towards low total payment prices. Therefore, we can both achieve differential privacy and a guaranteed approximation to the optimal payment, as will be proved in Section VI. Algorithm 1 normalizes $\exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big)$ and randomly picks a price p according to the following distribution (line 16) defined in Equation 11.

$$\Pr[p = x] = \frac{\exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big)}{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S(y)|}{2N c_{\max}}\Big)},\ \forall x \in \mathcal{P}. \quad (11)$$

After a price p is sampled, the winner set $\mathcal{S}$ is set to be the one corresponding to p, namely $S(p)$ (line 17). Finally, it returns the winner set $\mathcal{S}$ and the price p (line 18).

VI. Analysis

In this section, we provide formal theoretical analysis about the desirable properties of our DP-hSRC auction. First of all, we prove that the DP-hSRC auction is ϵ-differentially private in Theorem 2.

Theorem 2. *The DP-hSRC auction is ϵ-differentially private.*

Proof: We denote $\mathbf{b}$ and $\mathbf{b}'$ as two bid profiles that differ in only one worker's bid. $\forall x \in \mathcal{P}$, we have

$$\begin{aligned}
\frac{\Pr\big[M(\mathbf{b}) = x\big]}{\Pr\big[M(\mathbf{b}') = x\big]} &= \frac{\exp\Big(-\frac{\epsilon x |S(x)|}{2N c_{\max}}\Big)}{\exp\Big(-\frac{\epsilon x |S'(x)|}{2N c_{\max}}\Big)} \cdot \frac{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S'(y)|}{2N c_{\max}}\Big)}{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S(y)|}{2N c_{\max}}\Big)} \\
&\leq \exp\Big(\frac{\epsilon x N}{2N c_{\max}}\Big) \cdot \frac{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y (|S(y)| - N)}{2N c_{\max}}\Big)}{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S(y)|}{2N c_{\max}}\Big)} \\
&\leq \exp\Big(\frac{\epsilon}{2}\Big) \cdot \frac{\sum_{y \in \mathcal{P}} \exp\Big(\frac{-\epsilon y |S(y)| + \epsilon c_{\max} N}{2N c_{\max}}\Big)}{\sum_{y \in \mathcal{P}} \exp\Big(-\frac{\epsilon y |S(y)|}{2N c_{\max}}\Big)} \\
&= \exp\Big(\frac{\epsilon}{2}\Big) \cdot \exp\Big(\frac{\epsilon c_{\max} N}{2N c_{\max}}\Big) \\
&= \exp\Big(\frac{\epsilon}{2}\Big) \cdot \exp\Big(\frac{\epsilon}{2}\Big) \\
&= \exp(\epsilon).
\end{aligned}$$

That is,

$$\Pr\big[M(\mathbf{b}) = x\big] \leq \exp(\epsilon) \Pr\big[M(\mathbf{b}') = x\big],\ \forall x \in \mathcal{P}. \quad (12)$$

Therefore, we have $\Pr\big[M(\mathbf{b}) \in \mathcal{A}\big] \leq \exp(\epsilon) \Pr\big[M(\mathbf{b}') \in \mathcal{A}\big],\ \forall \mathcal{A} \subseteq \mathcal{P}$ and we arrive at the conclusion that the DP-hSRC auction is ϵ-differentially private. ■

We introduce the notation that $\Delta c = c_{\max} - c_{\min}$. Based on Theorem 2, we prove in Theorem 3 that the DP-hSRC auction is $\epsilon\Delta c$-truthful.

Theorem 3. *The DP-hSRC auction is $\epsilon\Delta c$-truthful.*

Proof: Similar to the proof of Theorem 2, we use $\mathbf{b}$ and $\mathbf{b}'$ to denote two bid profiles that differ in only one worker's bid. An equivalent form of Equation 12 proved in Theorem 2 is $\Pr\big[M(\mathbf{b}) = x\big] \geq \exp(-\epsilon)\Pr\big[M(\mathbf{b}') = x\big],\ \forall x \in \mathcal{P}$.

Therefore, the expectation of any worker w_i's utility taken over the output price distribution of the DP-hSRC auction mechanism $M(\cdot)$ given in Algorithm 1 satisfies that

$$\begin{aligned}
\mathbb{E}_{x\sim M(\mathbf{b})}\big[u_i(x)\big] &= \sum_{x\in\mathcal{P}} u_i(x)\Pr\big[M(\mathbf{b}) = x\big] \\
&\geq \sum_{x\in\mathcal{P}} u_i(x)\exp(-\epsilon)\Pr\big[M(\mathbf{b}') = x\big] \\
&= \exp(-\epsilon)\mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big] \\
&\geq (1-\epsilon)\mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big] \\
&= \mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big] - \epsilon\mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big].
\end{aligned}$$

Since the maximum price in $\mathcal{P}$ is $c_{\max}$ and the minimum possible cost for a worker is $c_{\min}$, we have that $u_i(x) \leq c_{\max} - c_{\min}, \forall x \in \mathcal{P}$. Therefore, we have $\mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big] \leq c_{\max} - c_{\min} = \Delta c$ and thus, $\mathbb{E}_{x\sim M(\mathbf{b})}\big[u_i(x)\big] \geq \mathbb{E}_{x\sim M(\mathbf{b}')}\big[u_i(x)\big] - \epsilon\Delta c$.

Therefore, we conclude that the DP-hSRC auction is $\epsilon\Delta c$-truthful. ■

Theorem 3 basically states that the proposed DP-hSRC auction upper bounds a worker's gain in her expected utility to bid untruthfully by $\epsilon\Delta c$. Therefore, we reasonably assume that each worker would bid truthfully in our DP-hSRC auction. Note that our DP-hSRC auction is $\epsilon\Delta c$-truthful in both the bidding bundle and price, namely any worker w_i bids her truthful bid $b_i^* = (\Gamma_i^*, c_i^*)$. In Theorem 4, we prove that our DP-hSRC auction is individual rational.

Theorem 4. *The DP-hSRC auction is individual rational.*

Proof: In every iteration of the main loop in Algorithm 1 (line 6-15), the candidate winner set $\mathcal{N}'$ is initialized as those workers whose bidding prices (i.e., ρ_k) are not larger than the given price $p = \rho_i$ (line 7). Furthermore, we have proved in Theorem 3 that every worker w_k bids truthfully, i.e., $\rho_k = c_k$. It means that for any given price p the winners are selected among the workers (i.e., w_k) such that $c_k \leq p$. As a consequence, any winner w_k's utility satisfies $u_k = p - c_k \geq 0$ and any loser's utility equals to 0. Therefore, we conclude that the DP-hSRC auction is individual rational. ■

Next, we provide our analysis about the algorithmic properties of the proposed DP-hSRC auction regarding the computational complexity and its approximation ratio to the optimal total payment in Theorem 5 and 6. Firstly, we analyze the computational complexity of our DP-hSRC auction in the following Theorem 5.

Theorem 5. *The computational complexity of the proposed DP-hSRC auction is $O(N^2K)$.*

Proof: The computational complexity of Algorithm 1 is dominated by the main loop (line 6-15), which terminates in worst case after N iterations. Furthermore, in every iteration of the inner loop (line 8-13), one worker is selected as a new winner. Thus, the inner loop also terminates in worst case after N iterations. Besides, within the inner loop, after a winner is selected the algorithm updates the Q'_j value for every task $\tau_j \in \mathcal{T}$ in the worst case. Therefore, the overall computational complexity of the DP-hSRC auction is $O(N^2K)$. ■

As proved in Theorem 5, our DP-hSRC auction described in Algorithm 1 has polynomial-time computational complexity depending on the number of workers N and the number of tasks K. Furthermore, the computational complexity provided in Theorem 5 does not depend on the cardinality of the possible price set $\mathcal{P}$, namely $|\mathcal{P}|$. Before we analyze the approximation ratio of the total payment generated by Algorithm 1 to the optimal total payment R_{OPT} in Theorem 6, we introduce Lemma 2 which is borrowed from [10] (Theorem 5 in [10]). We define the unit measure of every element in $\mathbf{q}$ and $\mathbf{Q}$ as Δq and introduce additionally the following two notations, i.e., $\beta = \max_{i:w_i\in\mathcal{N}} \sum_{j:\tau_j\in\Gamma_i} q_{i,j}$ and $m = \frac{1}{\Delta q}\sum_{j:\tau_j\in\mathcal{T}} Q_j$.

Lemma 2. *Given $\forall p \in \mathcal{P}$, we have that the cardinality of the winner set returned by the proposed DP-hSRC auction $S(p)$ and that of the minimum-cardinality winner set $S_{\text{OPT}}(p)$ satisfies that*

$$|S(p)| \leq 2\beta H_m |S_{\text{OPT}}(p)|. \tag{13}$$

The relationship between the cardinality of the two sets $S(p)$ and $S_{\text{OPT}}(p)$ given in Lemma 2 is an important intermediary result that will be utilized in the proof of the following Theorem 6, which shows the approximation ratio of the total payment generated by the DP-hSRC auction to the optimal total payment.

Theorem 6. *Suppose given any price $x \in \mathcal{P}$, Algorithm 1 gives us a total payment $R(x)$. Then, the expected total payment generated by the DP-hSRC auction denoted by $\mathbb{E}_{x\in\mathcal{P}}[R(x)]$ and the optimal payment R_{OPT} satisfies that*

$$\mathbb{E}_{x\in\mathcal{P}}\big[R(x)\big] \leq 2\beta H_m R_{\text{OPT}} + \frac{6Nc_{\max}}{\epsilon}\ln\left(e + \frac{\epsilon|\mathcal{P}|\beta H_m R_{\text{OPT}}}{c_{\min}}\right).$$

Proof: We use $R_{\min}$ and $R_{\max}$ to denote the minimum and maximum total payment generated by Algorithm 1 and we define the following sets $\mathcal{B}_t = \{x | R(x) > R_{\min} + t\}$, $\overline{\mathcal{B}}_t = \{x | R(x) \leq R_{\min} + t\}$ and $\mathcal{B}_{2t} = \{x | R(x) > R_{\min} +$

$2t\}$ for some constant $t > 0$. Then, we have

$$\Pr[x \in \mathcal{B}_{2t}] \leq \frac{\Pr[x \in \mathcal{B}_{2t}]}{\Pr[x \in \overline{\mathcal{B}}_t]} = \frac{\sum_{x \in \mathcal{B}_{2t}} \frac{\exp\left(-\frac{\epsilon R(x)}{2Nc_{\max}}\right)}{\sum_{y \in \mathcal{P}} \exp\left(-\frac{\epsilon R(y)}{2Nc_{\max}}\right)}}{\sum_{x \in \overline{\mathcal{B}}_t} \frac{\exp\left(-\frac{\epsilon R(x)}{2Nc_{\max}}\right)}{\sum_{y \in \mathcal{P}} \exp\left(-\frac{\epsilon R(y)}{2Nc_{\max}}\right)}}$$

$$= \frac{\sum_{x \in \mathcal{B}_{2t}} \exp\left(-\frac{\epsilon R(x)}{2Nc_{\max}}\right)}{\sum_{x \in \overline{\mathcal{B}}_t} \exp\left(-\frac{\epsilon R(x)}{2Nc_{\max}}\right)}$$

$$\leq \frac{|\mathcal{B}_{2t}| \exp\left(-\frac{\epsilon(R_{\min}+2t)}{2Nc_{\max}}\right)}{|\overline{\mathcal{B}}_t| \exp\left(-\frac{\epsilon(R_{\min}+t)}{2Nc_{\max}}\right)}$$

$$= \frac{|\mathcal{B}_{2t}|}{|\overline{\mathcal{B}}_t|} \exp\left(-\frac{\epsilon t}{2Nc_{\max}}\right).$$

Then, we can calculate $\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big]$ as follows.

$$\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big] = \sum_{x \in \overline{\mathcal{B}}_{2t}} R(x)\Pr[p = x] + \sum_{x \in \mathcal{B}_{2t}} R(x)\Pr[p = x]$$

$$\leq R_{\min} + 2t + R_{\max} \frac{|\mathcal{B}_{2t}|}{|\overline{\mathcal{B}}_t|} \exp\left(-\frac{\epsilon t}{2Nc_{\max}}\right)$$

$$\leq R_{\min} + 2t + R_{\max}|\mathcal{P}| \exp\left(-\frac{\epsilon t}{2Nc_{\max}}\right).$$

Therefore, for any $t \geq \ln\left(\frac{R_{\max}|\mathcal{P}|}{t}\right) \cdot \frac{2Nc_{\max}}{\epsilon}$, we have

$$\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big] \leq R_{\min} + 3t. \tag{14}$$

If we let $t = \ln\left(e + \frac{\epsilon|\mathcal{P}|R_{\max}}{2Nc_{\max}}\right) \cdot \frac{2Nc_{\max}}{\epsilon} \geq \frac{2Nc_{\max}}{\epsilon}$, we have

$$\ln\left(\frac{R_{\max}|\mathcal{P}|}{t}\right) \cdot \frac{2Nc_{\max}}{\epsilon} \leq \ln\left(e + \frac{R_{\max}|\mathcal{P}|\epsilon}{2Nc_{\max}}\right) \cdot \frac{2Nc_{\max}}{\epsilon} = t.$$

Therefore, we can simply let $t = \ln\left(e + \frac{\epsilon|\mathcal{P}|R_{\max}}{2Nc_{\max}}\right) \cdot \frac{2Nc_{\max}}{\epsilon}$ and substitute t into Equation 14. We have

$$\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big] \leq R_{\min} + \ln\left(e + \frac{\epsilon|\mathcal{P}|R_{\max}}{2Nc_{\max}}\right) \cdot \frac{6Nc_{\max}}{\epsilon}.$$

Furthermore, since $R_{\max} \leq \frac{c_{\max}}{c_{\min}} N R_{\min}$, we have

$$\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big] \leq R_{\min} + \ln\left(e + \frac{\epsilon|\mathcal{P}|R_{\min}}{2c_{\min}}\right) \cdot \frac{6Nc_{\max}}{\epsilon}.$$

Suppose the optimal total payment R_{OPT} is achieved when the price $p = p^*$, i.e., $R_{\text{OPT}} = p^*|S_{\text{OPT}}(p^*)|$. Then, we have

$$R_{\min} \leq p^*|S(p^*)| \leq 2\beta H_m p^*|S_{\text{OPT}}(p^*)| = 2\beta H_m R_{\text{OPT}}.$$

Finally, we arrive at the conclusion that

$$\mathbb{E}_{x \in \mathcal{P}}\big[R(x)\big] \leq 2\beta H_m R_{\text{OPT}} + \frac{6Nc_{\max}}{\epsilon} \ln\left(e + \frac{\epsilon|\mathcal{P}|\beta H_m R_{\text{OPT}}}{c_{\min}}\right)$$

and we finish the proof of Theorem 6. ■

VII. Performance Evaluation

In this section, we present the baseline methods that we use in the simulation, as well as the simulation settings and results.

A. Baseline Method

Firstly, we compare the expected total payment of the DP-hSRC auction with the optimal total payment R_{OPT}. Instead of solving the TPM problem approximately using the method in Algorithm 1 (line 6-15), the exact optimal solution $S_{\text{OPT}}(p)$ to the TPM problem given any fixed price $p \in \mathcal{P}$ is calculated. Then, the optimal total payment $R_{\text{OPT}} = \min_{p \in \mathcal{P}} p|S_{\text{OPT}}(p)|$ is derived by iterating over every possible price $p \in \mathcal{P}$.

Furthermore, we compare our DP-hSRC auction with a baseline auction mechanism. For any fixed price $p \in \mathcal{P}$, the baseline auction selects the workers in $\mathcal{N}' = \{w_i|\rho_i \leq p\}$ as winners according to the descending order of the value $\sum_{j:\tau_j \in \Gamma_i} q_{i,j}$ until the error bound constraints of all tasks are satisfied. Then, a price p is picked randomly using the same method in Algorithm 1 (line 16). It is easily verifiable that the baseline auction is also ϵ-differentially private, $\epsilon\Delta c$-truthful and individual rational.

B. Simulation Settings

Setting	ϵ	$c_{\min}$	$c_{\max}$	$\lvert\Gamma_i^*\rvert$	$\theta_{i,j}$	δ_j	N	K
I	0.1	10	60	$[10, 20]$	$[0.1, 0.9]$	$[0.1, 0.2]$	$[80, 140]$	30
II	0.1	10	60	$[10, 20]$	$[0.1, 0.9]$	$[0.1, 0.2]$	120	$[20, 50]$
III	0.1	10	60	$[50, 150]$	$[0.1, 0.9]$	$[0.1, 0.2]$	$[800, 1400]$	200
IV	0.1	10	60	$[50, 150]$	$[0.1, 0.9]$	$[0.1, 0.2]$	1000	$[200, 500]$

Table I
Simulation settings

In Table I, we present the simulation settings. In setting I, we fix the number of tasks as 30 and vary the number of workers from 80 to 140. The privacy budget ϵ is set to be 0.1 and $c_{\min}$ and $c_{\max}$ is 10 and 60 respectively. Every worker w_i's cost c_i^* for her interested bundle Γ_i^* is chosen uniformly at random from the numbers spaced at the interval of 0.1 in the range $[10, 60]$. $|\Gamma_i^*|$, $\theta_{i,j}$, and δ_j are generated uniformly at random from the intervals given in Table I. Furthermore, the price set $\mathcal{P}$ consists of all numbers spaced at the interval of 0.1 in the range $[35, 60]$. In setting II, we fix the number of workers as 120 and vary the number of tasks from 20 to 50. All the other parameters are the same as those in setting I. In setting III and IV, the parameter ϵ, $c_{\min}$, $c_{\max}$, $|\Gamma_i^*|$, $\theta_{i,j}$, δ_j, c_i^*, and $\mathcal{P}$ are generated using the same method as in the previous two settings. The difference is that we increase the input size of the settings. In setting III, we fix the number of tasks as 200 and vary the number of workers from 800 to 1400, whereas in setting IV, we fix the number of workers as 1000 and vary the number of tasks from 200 to 500. Moreover, all the optimal solutions to the TPM problem are calculated using the GUROBI optimization solver [46].

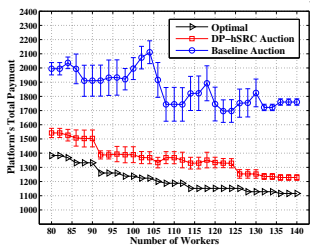


Figure 1. Platform's total payment under setting I

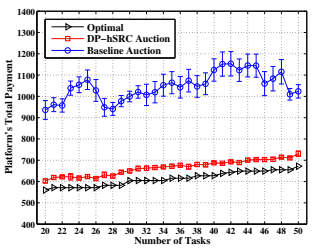


Figure 2. Platform's total payment under setting II

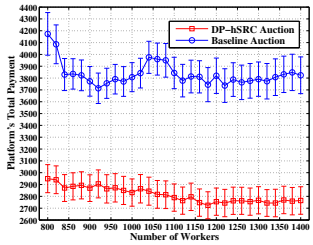


Figure 3. Platform's total payment under setting III

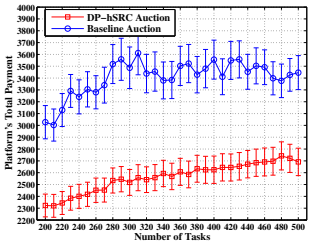


Figure 4. Platform's total payment under setting IV

C. Simulation Results

In Figure 1 and 2, for every given worker and task number, we sample a price from the price distribution derived by the DP-hSRC auction and the baseline auction, respectively, for 10000 times. The corresponding mean and standard deviation of the platform's total payment calculated using these price samples are plotted in Figure 1 and 2. From these two figures, we observe that the platform's average total payment of the DP-hSRC auction is far better than that of the baseline auction and fairly close to the optimal total payment R_{OPT}. Note that the nonsmoothness of the curves in Figure 1 and 2, as well as those in the forthcoming Figure 3 and 4, is due to the randomness in generating the problem instances.

N	80	88	96	104	112	120	128	136
DP-hSRC	0.156	0.158	0.157	0.161	0.161	0.156	0.165	0.159
Optimal	6.479	11.86	30.83	410.7	897.1	2337	2310	6139
K	20	24	28	32	36	40	44	48
DP-hSRC	0.152	0.153	0.153	0.158	0.157	0.157	0.160	0.162
Optimal	13.33	44.04	396.4	395.9	539.7	735.5	1188	2661

Table II
EXECUTION TIME (S) FOR SETTING I AND II

In Table II, we compare the execution time of the DP-hSRC auction and the algorithm that computes the optimal total payment R_{OPT}. From this table, we can observe that the DP-hSRC auction executes in significantly less time than the optimal algorithm. Furthermore, the execution time of the optimal algorithm becomes excessively long with large numbers of tasks and workers so that it is infeasible in practice. In contrast, regardless of the growth of the number of users and tasks, the DP-hSRC auction keeps low execution time. Hence, the DP-hSRC auction is much more computationally efficient than the optimal algorithm.

In Figure 3 and 4, we consider setting III and IV given in Table I. Setting III and IV have much more numbers of workers and tasks than setting I and II. Under setting III and IV, the scales of the problem have become so large that make it infeasible for the optimal algorithm to return the optimal results in reasonable time. In contrast, in Figure 3 and 4, we demonstrate that our DP-hSRC auction is still able to generate total payment far better than the baseline auction under setting III and IV.

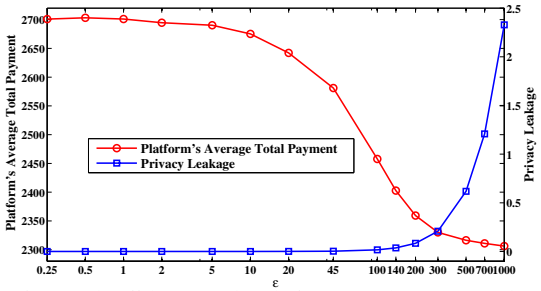


Figure 5. Trade-off between the platform's total payment and privacy leakage

In Figure 5, we plot the platform's average total payment and the privacy leackage of the DP-hSRC auction with the increasing of the privacy budget ϵ. For any fixed ϵ, we define the privacy leakage of the DP-hSRC auction as follows in Definition 8.

Definition 8 (Privacy Leakage)**.** *Suppose the two bid profiles* $\mathbf{b}$ *and* $\mathbf{b}'$ *that differ in only one worker's bid result in price distributions with probability mass functions (PMFs)* P *and* P'*. The privacy leakage of the two bid profiles is defined as the Kullback-Leibler (KL) divergence [47] of the two distributions represented as follows.*

$$\text{Privacy Leakage} = D_{\text{KL}}(P||P') = \sum_{x\in\mathcal{P}} P(x)\ln\left(\frac{P(x)}{P'(x)}\right).$$

The KL divergence captures the statistical difference of the two distributions P and P'. The larger the statistical difference is, the easier the two bid profiles $\mathbf{b}$ and $\mathbf{b}'$ will be distinguished and thus, the more the privacy leakage is. From Figure 5, we can observe that as the decreasing of ϵ, the privacy leakage decreases. Furthermore, such improvement in privacy protection comes at a cost of the increased total payment of the platform shown in Figure 5. Therefore, Figure 5 illustrates the trade-off between the platform's total payment and the privacy leakage of the DP-hSRC auction.

VIII. CONCLUSION

In this paper, motivated by the need for the protection of workers' privacy in MCS systems, we develop a differentially private incentive mechanism to incentivize worker participation without disclosing their sensitive bid information. The proposed mechanism is based on a novel design of single-minded reverse combinatorial auction with heterogeneous cost, and thus bears several advantageous properties

including approximate truthfulness, individual rationality, and computational efficiency. We conduct both theoretical analysis and extensive simulations to show that the proposed mechanism minimizes the expected total payment with a guaranteed approximation ratio to the optimal total payment.

REFERENCES

[1] R. K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T. F. Abdelzaher, "Greengps: A participatory sensing fuel-efficient maps application," in *MobiSys*, 2010.

[2] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li, "Jigsaw: Indoor floor plan reconstruction via mobile crowdsensing," in *Mobicom*, 2014.

[3] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "Aircloud: a cloud-based air-quality monitoring system for everyone," in *SenSys*, 2014.

[4] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification," *ACM Transactions on Sensor Networks (TOSN)*, vol. 11, no. 4, p. 55, 2015.

[5] L. Gao, F. Hou, and J. Huang, "Providing long-term participation incentive in participatory sensing," in *INFOCOM*, 2015.

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

[7] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, "Truthful incentive mechanisms for crowdsourcing," in *INFOCOM*, 2015.

[8] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Mobicom*, 2012.

[9] Z. Feng, Y. Zhu, Q. Zhang, L. Ni, and A. Vasilakos, "Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *INFOCOM*, 2014.

[10] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *MobiHoc*, 2015.

[11] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen, "Quality-driven auction based incentive mechanism for mobile crowd sensing," *Vehicular Technology, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[12] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *INFOCOM*, 2013.

[13] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *INFOCOM*, 2014.

[14] Z. Feng, Y. Zhu, Q. Zhang, H. Zhu, J. Yu, J. Cao, and L. Ni, "Towards truthful mechanisms for mobile crowdsourcing with dynamic smartphones," in *ICDCS*, 2014.

[15] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 12, pp. 3190–3200, Dec 2014.

[16] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *INFOCOM*, 2014.

[17] "Myheartmap," http://www.med.upenn.edu/myheartmap/.

[18] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and B. Hari, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *MobiSys*, 2008.

[19] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security*, 2011.

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

[21] R. Zhu, Z. Li, F. Wu, K. Shin, and G. Chen, "Differentially private spectrum auction with approximate revenue maximization," in *MobiHoc*, 2014.

[22] R. Zhu and K. Shin, "Differentially private and strategy-proof spectrum auction with approximate revenue maximization," in *INFOCOM*, 2015.

[23] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *INFOCOM*, 2012.

[24] M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in *MobiHoc*, 2015.

[25] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *MobiHoc*, 2015.

[26] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *ICDCS*, 2014.

[27] ——, "Providing privacy-aware incentives for mobile sensing," in *PerCom*, 2013.

[28] X. Niu, M. Li, Q. Chen, Q. Cao, and H. Wang, "Eppi: An e-cent-based privacy-preserving incentive mechanism for participatory sensing systems," in *IPCCC*, 2014.

[29] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *MobiHoc*, 2016.

[30] J. Sun and H. Ma, "Privacy-preserving verifiable incentive mechanism for online crowdsourcing markets," in *ICCCN*, 2014.

[31] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *SenSys*, 2008.

[32] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones," in *SenSys*, 2009.

[33] D. Oleson, A. Sorokin, G. P. Laughlin, V. Hester, J. Le, and L. Biewald, "Programmatic gold: Targeted and scalable quality assurance in crowdsourcing," in *HCOMP*, 2011.

[34] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *SIGMOD*, 2014.

[35] L. Su, Q. Li, S. Hu, S. Wang, J. Gao, H. Liu, T. Abdelzaher, J. Han, X. Liu, Y. Gao, and L. Kaplan, "Generalized decision aggregation in distributed sensing systems," in *RTSS*, 2014.

[36] Q. Li, Y. Li, J. Gao, L. Su, B. Zhao, M. Demirbas, W. Fan, and J. Han, "A confidence-aware approach for truth discovery on long-tail data," *Proc. VLDB Endow.*, vol. 8, no. 4, pp. 425–436, Dec. 2014.

[37] Y. Li, J. Gao, C. Meng, Q. Li, L. Su, B. Zhao, W. Fan, and J. Han, "A survey on truth discovery," *SIGKDD Explor. Newsl.*, 2016.

[38] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng, "Truth discovery on crowd sensing of correlated entities," in *SenSys*, 2015.

[39] H. Li, B. Zhao, and A. Fuxman, "The wisdom of minority: Discovering and targeting the right group of workers for crowdsourcing," in *WWW*, 2014.

[40] C.-J. Ho, S. Jabbari, and J. W. Vaughan, "Adaptive task assignment for crowdsourced classification," in *ICML*, 2013.

[41] L. Blumrosen and N. Nisan, "Combinatorial auctions," *Algorithmic Game Theory*, 2007.

[42] M. Babaioff, R. Lavi, and E. Pavlov, "Single-value combinatorial auctions and algorithmic implementation in undominated strategies," *J. ACM*, 2009.

[43] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar, "Differentially private combinatorial optimization," in *SODA*, 2010.

[44] A. Gopinathan and Z. Li, "A prior-free revenue maximizing auction for secondary spectrum access," in *INFOCOM*, 2011.

[45] N. Immorlica, A. R. Karlin, M. Mahdian, and K. Talwar, "Balloon popping with applications to ascending auctions," in *FOCS*, 2007.

[46] "Gurobi solver," http://www.gurobi.com/.

[47] S. Kullback and R. A. Leibler, "On information and sufficiency," *The annals of mathematical statistics*, pp. 79–86, 1951.