Crowd-Empowered Privacy-Preserving Data Aggregation for Mobile Crowdsensing

Lei Yang
CSE Department at UNR
leiy@unr.edu

Mengyuan Zhang
State Key Laboratory of Industrial Control Technology at ZJU
zhang418@zju.edu.cn

Shibo He
State Key Laboratory of Industrial Control Technology at ZJU
shibohe.cn@gmail.com

Ming Li
CSE Department at UNR
mingli@unr.edu

Junshan Zhang
School of ECEE at ASU
Junshan.Zhang@asu.edu

ABSTRACT

We develop an auction framework for privacy-preserving data aggregation in mobile crowdsensing, where the platform plays the role as an auctioneer to recruit workers for a sensing task. In this framework, the workers are allowed to report privacy-preserving versions of their data to protect their data privacy; and the platform selects workers based on their sensing capabilities, which aims to address the drawbacks of game-theoretic models that cannot ensure the accuracy level of the aggregated result, due to the existence of multiple Nash Equilibria. Observe that in this auction based framework, there exists externalities among workers' data privacy, because the data privacy of each worker depends on both her injected noise and the total noise in the aggregated result that is intimately related to which workers are selected to fulfill the task. To achieve a desirable accuracy level of the data aggregation in a cost-effective manner, we explicitly characterize the externalities, i.e., the impact of the noise added by each worker on both the data privacy and the accuracy of the aggregated result. Further, we explore the problem structure, characterize the hidden monotonicity property of the problem, and determine the critical bid of workers, which makes it possible to design a truthful, individually rational and computationally efficient incentive mechanism. The proposed incentive mechanism can recruit a set of workers to approximately minimize the cost of purchasing private sensing data from workers subject to the accuracy requirement of the aggregated result. We validate the proposed scheme through theoretical analysis as well as extensive simulations.

CCS CONCEPTS

• **Security and privacy → Privacy protections**; • **Networks → Network economics**; • **Human-centered computing → Mobile computing**; • **Theory of computation → Design and analysis of algorithms**;

Mobihoc '18, June 26–29, 2018, Los Angeles, CA, USA

ACM ISBN 978-1-4503-5770-8/18/06…$15.00
https://doi.org/10.1145/3209582.3209598

KEYWORDS

Crowd sensing, incentive mechanism, privacy-preserving, data aggregation

1 INTRODUCTION

1.1 Motivation

Mobile crowdsensing arises as a promising sensing paradigm that leverages the sensing capability of human-carried mobile devices to perform various sensing tasks (e.g., healthcare, environment monitoring, indoor localization, and smart transportation) [22]. By outsourcing the sensing task to the public crowd, mobile crowdsensing systems can collect fine-grained information effectively and efficiently. However, any individual involved in a sensing task inevitably authorizes the task agent a certain level of privilege to access her sensing data which can be sensitive, thereby giving rise to privacy concerns when being released to an untrusted party. This becomes a key challenge hindering individuals (workers) from participation, more than the consumption of the limited system resources (e.g., battery and computing power) of their mobile devices. Therefore, the success of mobile crowdsensing hinges upon the design of efficient incentive mechanisms to stimulate workers' participation.

Many incentive mechanisms have been developed for mobile crowdsensing systems (e.g., [1, 4, 7, 10–12, 14, 15, 21, 25–30]). Most of the existing works take into account only workers' sensing costs, and only a few recent works consider workers' privacy costs. However, in these works considering workers' privacy costs, either workers have no control of their data privacy (e.g., the platform is assumed to be trustworthy and fully responsible for protecting workers' private data in [15]), or the platform interacts with workers via game-theoretic models (e.g., [24]), which may end up with an inefficient equilibrium, i.e., the platform may not achieve a desirable accuracy level of the aggregated result. To address these issues, novel data aggregation for mobile crowdsensing is needed to allow not only each worker to protect their data privacy by themselves but also the platform to choose workers selectively based on their sensing quality to achieve a desirable accuracy level of the aggregated result. One possible solution to these issues is to allow workers to protect their data privacy by reporting noisy data [24]. Clearly, this approach would negatively impact the reliability of the

sensing results.[1] To ensure the accuracy of the aggregated results, the platform needs to devise more efficient incentive mechanisms that take into account workers' privacy protecting behaviors, in order to achieve a good balance between workers' data privacy and the accuracy of the aggregated results.

One key question is how to achieve a desirable accuracy level of the data aggregation in a cost effective manner when the workers report noisy data. As game-theoretic models cannot guarantee a desirable accuracy level of the data aggregation due to the existence of multiple Nash Equilibria (e.g., [24]), this paper employs an auction approach that takes into account the accuracy requirement when designing the incentive mechanism. However, allowing workers to add noise into their sensing results renders three major challenges for the incentive mechanism design:

- *Strategic Behavior.* When workers report noisy data, the platform does not know the true sensing data and the added noise. Thus, it is possible that workers could play strategically by adding more noise into their sensing data to enhance their data privacy during the data aggregation stage. Therefore, a new data aggregation scheme is required to allow the platform to control the noise level of workers' data without knowing their true sensing data. Moreover, workers may manipulate their bids to maximize their own benefits, which may lead to high costs of achieving a desirable accuracy level of the data aggregation. Therefore, a truthful incentive mechanism is required.
- *Externalities.* Compared with the existing works (e.g., [15]), where the platform adds noise into workers' sensing data and workers' data privacy depends on the noise added by the platform only, the data privacy of each worker in our paper depends on not only the noise added by herself but also the total noise in the aggregated result (see Section 2.3). In other words, the data privacy of each worker depends on which workers are selected to fulfill the task and how much noise the selected workers generate, which introduces *externalities*. This makes the design of incentive mechanism in this paper more challenging.
- *Computational Complexity.* To achieve a desirable accuracy level of the aggregated result in a cost-effective manner, the platform needs to find an optimal subset of workers to fulfill the sensing task. Because different workers have different valuations of their data privacy and workers' data privacy is interdependent due to externalities, it is of combinatorial nature to find an optimal subset of workers to minimize the system cost while achieving the desirable accuracy level. Therefore, a computationally efficient mechanism is needed.

1.2 Summary of Main Contributions

In this paper, we develop an auction framework for privacy-preserving data aggregation in mobile crowdsensing, where the workers submit their bids to the platform and the platform plays the role as an auctioneer to recruit workers for a sensing task. When aggregating noisy data from workers, the platform aims to minimize the cost of purchasing the *private* sensing data from the workers, while achieving a desirable accuracy level of the aggregated result. Our main contributions are summarized as follows:

- *Differentially Private Data Aggregation.* To tackle the challenge due to workers' strategic behaviors, we propose a differentially private data aggregation scheme by leveraging the celebrated concept of differential privacy. The key idea is to carefully design the noise distribution for each worker based on the divisible property of Laplace distribution, such that each worker can report a privacy-preserving version of their data based on the designed noise distribution and the platform can guarantee the differential privacy of each worker's data. By using this scheme, the platform can prevent workers from strategically adding large noise into workers' sensing data and control the noise level of their data without knowing their true sensing data.
- *Externalities.* Under the proposed differentially private data aggregation scheme, for different sets of workers, different noise distributions will be designed for the workers. In other words, the privacy of a worker would change if the platform chooses different workers, which introduces externalities. For the Laplace noise distribution, we explicitly characterize the externalities among workers and the impact of each worker's participation on the privacy of other workers, which is accounted for in the incentive mechanism design.
- *Privacy-Accuracy Tradeoff.* To maintain the accuracy of the aggregated result, the platform would reward workers more if the reported data is of higher accuracy (i.e., less noise is added). Clearly, there is a tradeoff between the (privacy) cost and the accuracy. We characterize the tradeoff between workers' data privacy and the accuracy of the aggregated result based on the concept of differential privacy. The accuracy of the aggregated result is characterized in terms of the distortion, due to the noise added by workers.
- *Differentially Private Data Auction.* Based on the proposed differentially private data aggregation, the design of the incentive mechanism boils down to solving a privacy auction of allocating the sensing task to a set of workers that can minimize the total payment to the workers, subject to the accuracy constraint of the aggregated result. We show that it is NP-hard to find the optimal solution to this problem. By exploring the problem structure, we discover the *hidden monotonicity* property of the problem and determine the critical bid of workers. Based on these findings, we propose a computationally efficient differentially private data auction scheme despite the combinatorial nature of the problem. Moreover, we show that the proposed differentially private data auction scheme is truthful, individually rational and close to the optimal solution. The performance of the proposed scheme is evaluated via extensive simulations.

1.3 Related Work

Incentive mechanism design for mobile crowdsensing systems has recently garnered much attention (e.g., [1, 4, 7, 11, 12, 14, 15, 21, 25–30]). Different models (e.g., auction [7, 14, 15, 25–30] and game-theoretic models [1, 4, 11, 12, 21]) have been utilized to design incentive mechanisms with different objectives, including social welfare maximization (e.g., [1, 8, 14]), cost or payment minimization (e.g., [7, 15]), and platform's profit maximization (e.g., [11, 27]). Most of the existing works (e.g., [7, 14, 25, 26, 28–30]) consider only the sensing costs of the participants.

[1]The reliability of the sensing results depends on the total noise added by the workers and the sensor quality of their mobile devices [15].

Recently, there has been much attention paid to data privacy (e.g., [2, 3, 9, 10, 15, 24, 27]). Most of these works (e.g., [2, 3, 9, 15, 27]) consider the case that the platform (i.e., the data collector) is trustworthy and the true data is reported to the platform, where workers have no control of their data privacy. Very recent works [23, 24] allow the workers to protect their data privacy by reporting noisy data and study how to trade private data in game-theoretic models, which, however, may end up with an inefficient equilibrium, i.e., the accuracy of the aggregated result cannot be guaranteed. To address these issues, this paper proposes a novel auction framework for mobile crowdsensing, where the workers can protect their data privacy by adding noise based on the noise distributions determined through the proposed data aggregation scheme and the platform can select the workers to minimize the cost of achieving a desirable accuracy level of the data aggregation. This differentiates our approach from the existing works on data aggregation in mobile crowdsensing.

The rest of the paper is organized as follows. In Section 2, we describe the crowd-empowered privacy-preserving data aggregation for mobile crowdsensing systems. In Section 3, we propose the incentive mechanism and analyze its properties. In Section 4, we evaluate the performance of the proposed incentive mechanism. The paper is concluded in Section 5.

2 PRIVACY-PRESERVING DATA AGGREGATION FOR MOBILE CROWDSENSING

2.1 System Overview

Consider a mobile crowdsensing system consisting of a centralized platform $\mathcal{A}$, a task agent $\mathcal{T}$ and a set of participating workers $\mathcal{N} \triangleq \{1, \cdots, N\}$, as illustrated in Fig. 1. The task requires workers to report to the platform their local sensing data of a specific object or phenomenon (e.g., spectrum sensing and environmental monitoring). To enhance the reliability of the result, the platform will aggregate the sensing data, as the reliability of each worker's sensing data may be different due to different sensor qualities [15]. *Different from the existing works on auctions in mobile crowdsensing systems (e.g., [1, 4, 7, 10–12, 14, 15, 21, 25–30]), we allow each individual to report a privacy-preserving version of her data to protect her own data privacy [24].*

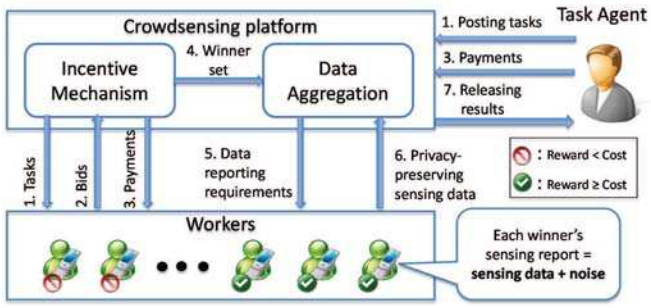

Figure 1: Framework of crowd-empowered privacy-preserving data aggregation.

Specifically, the workflow (see Fig. 1) of the proposed crowd-empowered privacy-preserving data aggregation is as follows:

- First, the task agent posts a task in the crowdsensing platform, which then announces the task to a set of N workers, denoted as $\mathcal{N}$ (step 1).
- **Incentive Mechanism.** Then, the workers submit their bids to the platform (step 2), where the bids reflect the valuation of privacy loss of each worker (see Section 2.2.1). Based on the incentive mechanism, the platform determines the winners, i.e., the workers to fulfill the task, and the corresponding payments to the winners (steps 3 and 4).
- **Data Aggregation.** Next, the platform sends the data reporting requirements to the winners and allows the winners to report a privacy-preserving version of their sensing data (steps 5 and 6).
- Finally, the platform releases the aggregated result to the task agent (step 7).

2.2 Crowdsensing Auction Model

In the crowdsensing system, the platform plays the role as an auctioneer who recruits workers to complete the sensing task and then aggregates the sensing data.

First, the platform (auctioneer) would elicit bids for unit privacy cost from the workers. The privacy cost of the workers is quantified using differential privacy (see Section 2.2.1). Let $\mathbf{b} = (b_1, \cdots, b_N)$ denote the vector of bids submitted by the workers and $\mathbf{b}_{-i}$ denote the bid vector without worker i's bid. Without loss of generality, we assume that workers' bids are ordered in the increasing order, i.e., $b_1 \leq b_2 \leq \cdots \leq b_N$. To prevent manipulations of bids that may lead to system performance degradation (e.g., high costs), a truthful incentive mechanism is required, which is discussed in Section 3. Moreover, as the platform does not know the noise added by workers, workers can play strategically by adding more noise into their sensing data to enhance their data privacy during the data aggregation stage. Therefore, a novel data aggregation scheme is required, which is discussed in Section 2.3.

Then, the auction outputs an allocation result $(\mathbf{x}, \mathbf{p})$, in which $\mathbf{x} = (x_1, \cdots, x_N)$ indicates the participants and $\mathbf{p} = (p_1, \cdots, p_N)$ indicates the amount of payments to the participants. Specifically, $x_i \in \{0, 1\}$ denotes if worker i is selected to execute the task: $x_i = 1$ means that worker i is selected (i.e., winner) and $x_i = 0$, otherwise. Accordingly, we define $\mathcal{S}$ as the winner set with S workers. For each worker, the platform will pay p_i amount of reward to worker $i \in \mathcal{N}$ to collect the private data from worker i, and use the data in a differentially private manner after the data aggregation (see Section 2.3).

2.2.1 Privacy Cost. According to the utility theoretic characterization of differential privacy [9], the privacy cost can be modeled as the difference between the utility with true data vector and the utility with perturbed data vector, which is a linear function of worker's privacy ϵ_i. Let $v_i > 0$ denote worker i's intrinsic valuation of unit privacy cost. The privacy cost of worker i can be given by

$$c_i = v_i \epsilon_i(\mathbf{x}). \tag{1}$$

Intuitively, a larger value of v_i indicates that worker i has a higher valuation of privacy loss by revealing her sensing data. We assume that all unit privacy costs v_i are unknown to the platform or to the

other workers. *Note that this cost function has been used in many existing works (e.g., [3, 9, 15, 27]). However, the worker's privacy ϵ_i in (1) in this paper is a function of* $\mathbf{x}$ *and depends on not only the noise added by herself but also the total noise in the aggregated result, which introduces externalities (see Section 2.3). This is one major difference between this work and other related works in mobile crowdsensing (e.g., [15]), where the privacy cost of a worker purely depends on her own participation.*

2.2.2 Worker's Utility. In the crowdsensing system, workers are assumed to be selfish and strategic, in order to maximize their own utilities. Based on the privacy cost (1), the utility u_i of worker i can be given as,

$$u_i(b_i, \mathbf{b}_{-i}) = p_i(b_i, \mathbf{b}_{-i}) - c_i = p_i(b_i, \mathbf{b}_{-i}) - v_i\epsilon_i(\mathbf{x}), \quad (2)$$

where u_i and p_i are functions of $\mathbf{b}$, given ϵ_i and $\mathbf{x}$. Here it holds that for a non-participant $i \in \mathcal{N}$ (i.e., $x_i = \epsilon_i = p_i = 0$), her utility turns out to be zero. Notice that we do not explicitly include the sensing cost of carrying out the task into the utility function (2) in order to ease the presentation. Meanwhile, our results in this paper can be easily extended to incorporate the sensing cost as in [5, 12]. For example, similar to [12], letting s_i denote the sensing cost of user i, we can modify the individual utility of user i as $u_i = p_i - s_i - \epsilon_i v_i$, and define $p_i' = p_i - s_i$ to incorporate the sensing cost in the reward. Therefore, our results can be extended to this case.

2.2.3 Design Objectives. We aim to design an auction based allocation mechanism that minimizes the total payment to the workers with satisfactory data aggregation accuracy, by designing an incentive mechanism (see Section 3) with the following desirable properties:

- **Truthfulness:** Each worker i can maximize her utility by truthfully bidding her privacy valuation, i.e., $u_i(v_i, \mathbf{b}_{-i}) \geq u_i(b_i, \mathbf{b}_{-i})$ for any $\mathbf{b}$.
- **Individual Rationality:** Each worker i can obtain a non-negative utility, i.e., $u_i = p_i - c_i \geq 0$.
- **Cost Minimization:** The mechanism can minimize the total payment to the workers.
- **Computational Efficiency:** The solution $(\mathbf{x}, \mathbf{p})$ can be computed in polynomial time.

2.3 Differentially Private Data Aggregation

When aggregating the data, each winner i will report a privacy-preserving version $\hat{d}_i$ of her data d_i by adding random noise n_i. Without loss of generality, we assume that all the sensing data d_i are normalized values within the range $[0, 1]$. In this paper, we consider a weighted aggregation operation f to calculate the aggregated result r based on workers' data. Let $\mathbf{d}$ be the vector of workers' data. The aggregated result r can be written as

$$r = f(\mathbf{d}) = \sum_{i\in\mathcal{N}} w_i(d_i + n_i)x_i = \sum_{i\in\mathcal{S}} w_i(d_i + n_i), \quad (3)$$

where $w_i > 0$ is the normalized weight of worker i such that the sum of these weights is equal to 1. Similar to [15, 17, 19], the weighted aggregation is to capture the effect of workers' diverse skill levels on the calculation of the aggregated results. Intuitively, higher weights will be assigned to workers whose sensing data are more likely to be close to the ground truths. This makes the aggregated results closer to the data provided by more reliable workers, which have been used by many state-of-the-art data aggregation methods [15, 17, 19]. The choice of weights can be based on workers' skill levels as in [15], which is *a priori* known to the platform and the workers.

In this paper, we quantify the privacy loss incurred in data aggregation based on the celebrated concept of differential privacy [6], and the proposed differentially private data aggregation is defined as follows.

Definition 1 (Differentially Private Data Aggregation). *An aggregation operation $f : [0, 1]^S \to \mathbb{R}$ is ϵ_i-differentially private with respect to worker i, if for any pair of neighboring vectors $\mathbf{d}$ and $\mathbf{d}_{(i)}$ differing only in the i^{th} worker's data and any set of aggregation results $O \subseteq Range(f)$, the following inequality holds:*

$$Pr[f(\mathbf{d}) \in O] \leq \exp(\epsilon_i) Pr[f(\mathbf{d}_{(i)}) \in O], \quad (4)$$

with ϵ_i being a positive parameter.

It follows that worker i's data is used in an ϵ_i-differentially private manner under operation f. This definition differs slightly from the definition in [6], which is stated in terms of the worst case privacy (i.e., ϵ-differentially private, where $\epsilon = \sup_i \epsilon_i$).

Given the aggregation operation f, a well-known method to provide differential privacy is to add random noise drawn from a Laplace distribution to this function [6]. As we allow each worker to add noise by themselves, we need to carefully design the noise distribution for each worker such that the sum of these noise is equivalent to the random noise drawn from a Laplace distribution, i.e., the aggregated noise $n = \sum_{i\in\mathcal{S}} w_i n_i$ follows the Laplace distribution.

Proposition 1. *For the aggregation operation f in (3), define $\epsilon_i = s_i(f)/\sigma$, where $s_i(f) = \max_{\mathbf{d}, \mathbf{d}_{(i)}\in[0,1]^S} |f(\mathbf{d}) - f(\mathbf{d}_{(i)})|$ is the sensitivity of f to the i^{th} entry d_i and σ is the parameter of the Laplace distribution. The aggregation operation f is ϵ_i-differentially private with respect to worker i, if $n_i = G_1(S, \sigma/w_i) - G_2(S, \sigma/w_i)$ for all $i \in \mathcal{S}$ are independent, where $G_1(S, \sigma/w_i)$ and $G_2(S, \sigma/w_i)$ are i.i.d. random variables following gamma distribution with pdf $g(x; S, \sigma/w_i) = \frac{1}{\Gamma(1/S)}(\frac{w_i}{\sigma})^{\frac{1}{S}} x^{\frac{1}{S}-1} e^{\frac{w_i x}{\sigma}}$.*

Proof. To show Proposition 1, it suffices to show that the aggregated noise follows the Laplace distribution. Based on the divisible property of Laplace distribution [16], the Laplace distribution is divisible and can be constructed as the sum of *i.i.d.* gamma distributions. Based on the scaling law of gamma distribution, $w_i n_i = G_1(S, \sigma) - G_2(S, \sigma)$. Therefore, we have

$$\sum_{i\in\mathcal{S}} w_i n_i = \sum_{i\in\mathcal{S}} (G_1(S, \sigma) - G_2(S, \sigma)) = L(\sigma), \quad (5)$$

where the second equality follows from the divisible property of Laplace distribution [16], which concludes the proof. □

Based on Proposition 1, if the noise distribution of each worker is carefully designed, the aggregation operation f in (3) is ϵ_i-differentially private with respect to worker i. Therefore, we propose the data aggregation mechanism in Algorithm 1. In Algorithm 1, the platform needs to inform the workers only the values of S and σ/w_i, based on which each worker generates a random noise and reports $\hat{d}_i$ back to the platform.

Remarks:

Algorithm 1 Differentially Private Data Aggregation

1: **Input:** Worker set $\mathcal{S}$, Number of workers S, weight of each worker $w_i, \forall i \in \mathcal{S}$, Laplace distribution parameter σ
2: **Output:** Aggregated result r.
3: For each worker $i \in \mathcal{S}$, the platform informs the values of parameters S and $\frac{\sigma}{w_i}$.
4: Each worker generates a random noise n_i based on the distribution of $G_1(S, \sigma/w_i) - G_2(S, \sigma/w_i)$, and then reports $\hat{d}_i = d_i + n_i$ to the platform.
5: The platform aggregates the data from the workers using (3) and releases the aggregated result r to the task agent.

- Note that in the proposed data aggregation algorithm, the platform does not know the true value of worker's data, but a privacy-preserving version of her data, which is generated using the noise distributions that the workers agree with in the proposed auction framework. By doing so, the proposed algorithm can not only allow the workers to report noisy data to protect their privacy but also prevent workers from strategically adding large noise into their sensing data, as it is easy to check if the distribution of each worker's reports follows the assigned noise distribution.
- Note that for different sets of winners, different noise distributions will be assigned to the winners. In other words, the privacy of each winner depends on the selection of the winner set, which introduces the *externalities*. This makes the design of incentive mechanism in this paper different from the existing works on auctions in mobile crowdsensing systems.

2.4 Privacy versus Accuracy

When allowing workers to report noisy data, the noise added into the aggregated result would inevitably reduce the accuracy of the result. From Proposition 1, we observe that ϵ_i depends on the value of σ. The higher the value of σ, the smaller ϵ_i, and hence, the better the privacy guarantee. However, the higher the value of σ, the lower the accuracy of the aggregated result. *Clearly, there is a natural trade-off between workers' data privacy and the accuracy of the aggregated result.*

To characterize the accuracy, we introduce the notion of *distortion* between two aggregation functions: one using all the workers' data with no noise and the other using the selected workers' data with noise (i.e., the aggregated result r in (3)). As the platform needs to pay for the workers' data, it would be costly to get all workers' data and workers would also add noise to protect their data privacy. Therefore, we can treat the aggregation of all the workers' data with no noise as the benchmark.

Definition 2 (Distortion). *Given the vector* $\mathbf{x}$*, the distortion* $\delta(\mathbf{x})$ *is defined as*

$$\delta(\mathbf{x}) = \max_{\mathbf{d} \in [0,1]^N} \mathbb{E}[(\sum_{i \in \mathcal{N}} w_i d_i - \sum_{i \in \mathcal{N}} w_i (d_i + n_i) x_i)^2]. \quad (6)$$

In Definition 2, the distortion is defined as the maximum of expected deviation from the true result for any sensing data reported by the workers. It is clear that the distortion depends on the set of workers fulfilling the task and the noise added into the data. Their dependence is quantified by the following proposition.

Proposition 2 (Privacy versus Distortion). *Given* x_i *and* w_i *for all the workers, under the aggregation function (3), the privacy of each worker and the distortion of the aggregated result can be given as*

$$\epsilon_i = \frac{w_i x_i}{\sigma}, \forall i \in \mathcal{N} \quad (7)$$

$$\delta(\mathbf{x}) = (\textstyle\sum_{i \in \mathcal{N}} w_i (1 - x_i))^2 + 2\sigma^2. \quad (8)$$

Proof. Given x_i and w_i under the aggregation function (3), we have

$$s_i(f) = \max_{\mathbf{d}, \mathbf{d}_{(i)} \in [0,1]^S} |w_i (d_i - d_i') x_i| = w_i x_i.$$

Therefore, we have $\epsilon_i = \frac{s_i(f)}{\sigma} = \frac{w_i x_i}{\sigma}$. For the distortion, we have

$$\begin{aligned}
\delta(\mathbf{x}) &= \max_{\mathbf{d} \in [0,1]^N} \mathbb{E}[(\sum_{i \in \mathcal{N}} w_i d_i - \sum_{i \in \mathcal{N}} w_i (d_i + n_i) x_i)^2] \\
&\overset{(a)}{=} \max_{\mathbf{d} \in [0,1]^N} \mathbb{E}[(\sum_{i \in \mathcal{N}} w_i d_i (1 - x_i) - \sum_{i \in \mathcal{S}} w_i n_i)^2] \\
&\overset{(b)}{=} \max_{\mathbf{d} \in [0,1]^N} (\sum_{i \in \mathcal{N}} w_i d_i (1 - x_i))^2 + 2\sigma^2 \\
&= (\sum_{i \in \mathcal{N}} w_i (1 - x_i))^2 + 2\sigma^2,
\end{aligned}$$

where (a) follows from equation (3) that $\sum_{i \in \mathcal{N}} w_i n_i x_i = \sum_{i \in \mathcal{S}} w_i n_i$, and (b) follows from Proposition 1 that $\sum_{i \in \mathcal{S}} w_i n_i$ is a Laplace random variable with zero mean and $2\sigma^2$ variance. □

From Proposition 2, it is clear that given σ, the more workers fulfilling the task, the less the distortion; given the set of selected workers $\mathcal{S}$, the higher the value of σ, the smaller ϵ_i (i.e., better privacy) and the worse the distortion. Following [3], we call the aggregated operation f in (3) *canonical* if the Laplace noise added by workers has a parameter of the following form

$$\sigma = \sigma(\mathbf{x}) = \textstyle\sum_{i \in \mathcal{N}} w_i (1 - x_i). \quad (9)$$

Based on (9), the privacy of each worker and the distortion of the aggregated result can be given as

$$\epsilon_i(\mathbf{x}) = \frac{w_i x_i}{\sum_{i \in \mathcal{N}} w_i (1 - x_i)}, \forall i \in \mathcal{N} \quad (10)$$

$$\delta(\mathbf{x}) = 3(\textstyle\sum_{i \in \mathcal{N}} w_i (1 - x_i))^2. \quad (11)$$

Eqs. (10) and (11) introduce *externalities* among the workers such that the data privacy of worker i depends on other workers' participations. Specifically, the more participants, the less the distortion but the larger ϵ_i (i.e., worse privacy). Intuitively, as the same sensing task is fulfilled by all the workers, the more participants, the more easily the true data can be figured out (i.e., the more privacy loss). Moreover, we need to carefully choose the workers as they have different skill levels (i.e., w_i) that may contribute differently to the distortion. Further, the costs of choosing different workers are different. Therefore, it is a challenging task to find a suitable set of workers to fulfill the sensing task.

3 INCENTIVE MECHANISM

3.1 Mathematical Formulation

The goal of crowdsensing auction is to minimize the total payment to the workers such that the accuracy of the aggregated result is above certain predetermined threshold (in other words, the distortion is below a threshold Δ). Specifically, this problem can be

formulated as

$$\begin{array}{ll} \text{minimize} & \sum_{i\in\mathcal{N}} p_i \\ \text{subject to} & p_i \geq b_i\epsilon_i(\mathbf{x}),\ \forall i \in \mathcal{N},\ \text{(Individual rationality)} \\ & \delta(\mathbf{x}) \leq \Delta,\ \text{(Accuracy requirement)} \\ & x_i \in \{0,1\}, \forall i \in \mathcal{N}. \end{array} \tag{12}$$

In problem (12), the constraints of individual rationality ensure that each worker can obtain non-negative utility. For the accuracy requirement constraint, the threshold will generally determine the total payment and the privacy protection levels of the workers. With a low threshold (i.e., high accuracy), the platform would pay more to the workers to obtain less noisy data (i.e., worse privacy for the workers). Note that different from most works on crowdsensing, problem (12) considers the externalities among workers such that workers' data privacy depends on each other, which has been discussed in Sections 2.3 and 2.4. Due to the externalities, designing an incentive mechanism to solve (12) is a challenging task. Theorem 1 shows that problem (12) is NP-hard.

Theorem 1. *The crowdsensing auction problem (12) is NP-hard.*

To show Theorem 1, we first establish the equivalence between problem (12) and the following problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i\in\mathcal{N}} b_i\epsilon_i(\mathbf{x}) \\ \text{subject to} & \sum_{i\in\mathcal{N}} w_i x_i \geq W, \\ & x_i \in \{0,1\}, \forall i \in \mathcal{N}, \end{array} \tag{13}$$

where $W = \sum_{i\in\mathcal{N}} w_i - (\Delta/3)^{1/2}$.

Lemma 1. *The optimal allocation $\mathbf{x}^*$ for problem (12) is the same as that for problem (13).*

Proof. Observe that to minimize (12), p_i is always equal to $b_i\epsilon_i(\mathbf{x}^*)$. Therefore, the inequalities for individual rationality are tight. In other words, minimizing $\sum_{i\in\mathcal{N}} p_i$ is equivalent to minimizing $\sum_{i\in\mathcal{N}} b_i\epsilon_i(\mathbf{x})$. Next, we can rewrite the constraint $\delta(\mathbf{x}) \leq \Delta$ as $\sum_{i\in\mathcal{N}} w_i x_i \geq W$ after some algebra, which concludes the proof. □

It is easy to show that problem (13) is reducible to a reverse binary knapsack problem, which is NP-hard. Based on Lemma 1, Theorem 1 follows.

3.2 Mechanism Design

From Theorem 1, problem (12) is computationally hard when the cardinality of $\mathcal{N}$ is large. To tackle this challenge, we propose a computationally efficient mechanism (see Algorithms 2 and 3), namely *differentially private data auction (DPDA)*, which is truthful and individually rational and can find the set of winners close to the optimal allocation $\mathbf{x}^*$ for problem (12), as discussed in Section 3.3.

In Algorithm 2, the idea is to first find the solution C of the fractional relaxation of problem (13), i.e.,

$$\begin{array}{ll} \text{minimize} & \sum_{i\in\mathcal{N}} b_i\epsilon_i(\mathbf{x}) \\ \text{subject to} & \sum_{i\in\mathcal{N}} w_i x_i \geq W, \\ & 0 \leq x_i \leq 1, \forall i \in \mathcal{N}, \end{array} \tag{14}$$

which is chosen as the target cost. Based on the target cost C, the set of winners can be determined by choosing the smallest set of workers with the total cost greater than or equal to C, because problem (14) is less constrained than problem (13) and thus C is a lower bound of the solution to problem (13). To find this smallest set of workers, we explore the solution structure of problem (14). Based on the relationship between problem (14) and problem (13), we discover the property of *monotonicity* (see the proof of Theorem 4 in Appendix), based on which the set of winners can be found by gradually adding the workers into the winner set until the total cost is greater than or equal to the target cost (see the main loop (line 6-10) in Algorithm 2). Essentially, we want to find the smallest k such that $\sum_{i\leq k} b_i w_i/(\sum_{i\geq k+1} w_i) \geq C$, i.e., $k = \min\{j : \sum_{i\leq j} b_i w_i/(\sum_{i\geq j+1} w_i) \geq C,\ \forall j \in \mathcal{N}\}$, and all the workers with $i \leq k$ are in the winner set. Note that due to the externalities, this monotonicity property is *hidden* in problem (13), which makes our problem more technically challenging than the existing auction works on mobile crowdsensing.

In Algorithm 3, we leverage the critical value approach in auction theory [20]. The idea is to determine the critical bid b_c such that a worker will not be selected if her bid is larger than or equal to b_c. Specifically, we first remove worker i from the worker set $\mathcal{N}$ and determine the smallest bid that makes a worker not a winner, which is done for each worker (line 5 in Algorithm 3). Note that the bids are ordered in the increasing order. The critical bid is determined based on the supremum of all these bids (line 6 in Algorithm 3). Using this critical bid, we determines the payment for each winner based on their weights (line 8 in Algorithm 3). From the analysis of DPDA in Section 3.3, we can see that the solution given by Algorithms 2 and 3 is feasible and close to the optimal solution to problem (12).

For the complexity of Algorithm 2, we need to solve C for problem (14), which is a linear fractional program. To efficiently solve C, we can transform problem (14) into a linear program based on the following lemma.

Lemma 2. *Problem (14) is equivalent to the following linear program:*

$$\begin{array}{ll} \textit{minimize} & \sum_{i\in\mathcal{N}} b_i w_i y_i \\ \textit{subject to} & \sum_{i\in\mathcal{N}} w_i y_i \geq W z, \\ & 0 \leq y_i \leq z, \forall i \in \mathcal{N}, \\ & \sum_{i\in\mathcal{N}} w_i z - \sum_{i\in\mathcal{N}} w_i y_i = 1. \end{array} \tag{15}$$

Proof. To show the equivalence, we will show that any feasible point in problem (14) is also feasible in problem (15) with the same objective value and vice versa. We note that if $\mathbf{x}$ is feasible in problem (14), then $y_i = \frac{x_i}{\sum_{i\in\mathcal{N}} w_i(1-x_i)}$, $\forall i \in \mathcal{N}$ and $z = \frac{1}{\sum_{i\in\mathcal{N}} w_i(1-x_i)}$ are feasible in problem (15), with the same objective value $\sum_{i\in\mathcal{N}} b_i w_i y_i = \sum_{i\in\mathcal{N}} b_i\epsilon_i(\mathbf{x})$. It follows that the optimal value of problem (14) is greater than or equal to the optimal value of problem (15). Conversely, note that $z > 0$ in problem (15). If y_i and z are feasible in problem (15), then $x_i = y_i/z$ is feasible in problem (14) with the same objective value $\sum_{i\in\mathcal{N}} b_i\epsilon_i(\mathbf{x}) = \sum_{i\in\mathcal{N}} b_i w_i y_i$. Therefore, the optimal value of problem (14) is less than or equal to the optimal value of problem (15). Therefore, problem (14) is equivalent to problem (15). □

Based on Lemma 2, we can solve C by solving a linear program (15). Note that the computational complexity of Algorithm 2 consists of two parts: solving a linear program (15) (line 3) and finding the set of winners (line 6-10). To solve (15) efficiently, we can use many solvers for linear programs, e.g., CPLEX [13], which can solve

the linear program (15) in polynomial time [18]. To find the set of winners, it takes at most $O(N)$ time in the worst case. Therefore, Algorithm 2 can determine the winner set for problem (12) in polynomial time. For Algorithm 3, it needs to run Algorithm 2 for each winner, and the worst case is to run N times, which means that it is also solvable in polynomial time.

Algorithm 2 Differentially Private Data Auction: Winner Determination

Input: worker set $\mathcal{N}$, weight of each worker $w_i, \forall i \in \mathcal{N}$, bid of each worker $b_i, \forall i \in \mathcal{N}$.
Output: winner set $\mathcal{S}$.
Find the target cost C by solving problem (15).
Let $k = 1, x_1 = 1$ and $x_i = 0,\ \forall i = 2, \ldots, N$.
Set $\mathcal{S} = \{1\}$ and compute $C' = b_1\epsilon_1(\mathbf{x})$.
while $C' < C$ **do** \\ Find the set of winners
  $k = k + 1$.
  Set $x_k = 1$ and $\mathcal{S} = \mathcal{S} \cup \{k\}$.
  $C' = \sum_{i=1}^{k} b_i\epsilon_i(\mathbf{x})$.
end while
return $\mathcal{S}$.

Algorithm 3 Differentially Private Data Auction: Payment Determination

Input: worker set $\mathcal{N}$, weight of each worker $w_i, \forall i \in \mathcal{N}$, bid of each worker $b_i, \forall i \in \mathcal{N}$, winner set $\mathcal{S}$.
Output: payments $\mathbf{p}$.
Set $\mathbf{p} = (0, \ldots, 0)$ and $b_c = b_{k+1}$, where k is the worker's index in $\mathcal{S}$ with the largest bid.
for each $i \in \mathcal{S}$ **do** \\ Find the critical bid
  Run Algorithm 2 on $\mathcal{N} \setminus \{i\}$ to get the winner set $\mathcal{S}'$ with k' being the worker's index in $\mathcal{S}'$ with the largest bid.
  $b_c = \min\{b_c, b_{k'+1}\}$.
end for
For each $i \in \mathcal{S}$, $p_i = \frac{b_c w_i}{\sum_{i\in\mathcal{N}\setminus\mathcal{S}} w_i}$.
return $\mathbf{p}$.

3.3 Analysis of DPDA

In this section, we will prove that DPDA is truthful, individually rational, and α-approximation with respect to the optimal cost.

First, we analyze the truthfulness of DPDA.

Theorem 2. *DPDA is truthful.*

Proof. To show DPDA is truthful, it is sufficient to show that users cannot improve their utilities by deviating their bids from their true valuations. Note that in DPDA, the winner is determined by the ranking of her bid in the set $\mathcal{N}$ and the higher the ranking, the lower the chance of being selected. Moreover, the critical bid determined by Algorithm 3 does not depend on the value of winners' bids. In what follows, we discuss the cases with an untruthful bid $\tilde{b}_i$ of worker i.

- Overbidding $\tilde{b}_i > v_i$. In this case, the ranking of worker i may move backward. If she could win the auction by truthfully bidding v_i and she remains in the winner set by overbidding, then her utility will remain the same because the critical bid b_c determined by Algorithm 3 will remain the same; if she loses the auction by overbidding, her utility will be zero. If she loses the auction by truthfully bidding, then she will still lose by overbidding. In either cases, worker i cannot improve her utility.
- Underbidding $\tilde{b}_i < v_i$. In this case, the ranking of worker i may move forward in the group. If she could win the auction by truthfully bidding v_i, then her utility cannot be improved since she must still remains in the winner set and the critical bid remains the same. If she loses the auction by truthfully bidding but underbidding helps her become a winner, her utility would be $u_i = \frac{(b_c - v_i)w_i}{\sum_{i\in\mathcal{N}\setminus\mathcal{S}} w_i}$. Since she is not originally in the winner set, it means that $v_i \geq b_c$, which leads to her utility $u_i \leq 0$.

Therefore, DPDA is truthful. □

Next, we analyze the individual rationality of DPDA.

Theorem 3. *DPDA is individually rational.*

Proof. For each worker in the winner set, we have

$$p_i = \frac{b_c w_i}{\sum_{i\in\mathcal{N}\setminus\mathcal{S}} w_i} \geq \frac{b_i w_i}{\sum_{i\in\mathcal{N}\setminus\mathcal{S}} w_i} = c_i,$$

since $b_c \geq b_i,\ \forall i \in \mathcal{S}$. For all workers who lose the auction, $p_i - c_i = 0$. Therefore, we have $p_i - c_i \geq 0$ for all the workers, i.e., DPDA is individually rational. □

Then, we analyze the approximation ratio of DPDA. The idea is to first characterize the optimal solution to problem (14), which, however, is still challenging, due to the externalities. To tackle this challenge, we explore the structure of problem (14) and discover the *hidden monotonicity* property after transforming problem (14) into an equivalent problem. Based on this finding, we show that DPDA satisfies the accuracy requirement of problem (12) and derive the approximation ratio of DPDA by using the relationship between the outputs of DPDA and the optimal solution to problem (12). The results are summarized in the following theorem.

Theorem 4. *DPDA satisfies the accuracy requirement (i.e., $\delta(\mathbf{x}) \leq \Delta$) and is α-approximation with respect to the optimal cost, where $\alpha = \frac{(b_k + C)w_k}{C\sum_{i\geq k} w_i - \sum_{i\leq k-1} b_i w_i} \geq 1$.*

The proof of Theorem 4 is given in Appendix.

4 PERFORMANCE EVALUATION

4.1 Simulation Setup

In our simulation, the bids are generated uniformly at random from the interval $[1, 20]$ and the weights are first generated uniformly at random from the interval $[1, 10]$ and then normalized. The number of workers N varies from 200 to 400. The distortion is normalized by some largest distortion $\Delta_{\max}$ such that W is always positive under different distortions. The optimal solutions to the problem (13) are calculated based on the bisection algorithm using the CPLEX optimization solver [13]. To the best of our knowledge, as there are no auction mechanisms for mobile crowdsensing allowing workers to report noisy data while considering the externalities, we examine

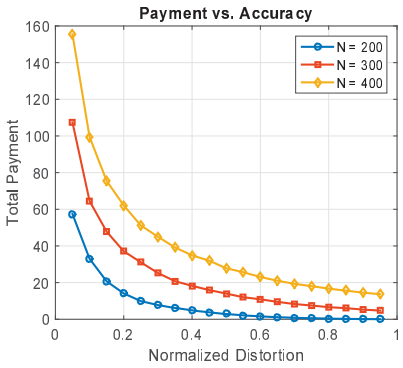

Figure 2: Payments under different accuracy requirements.

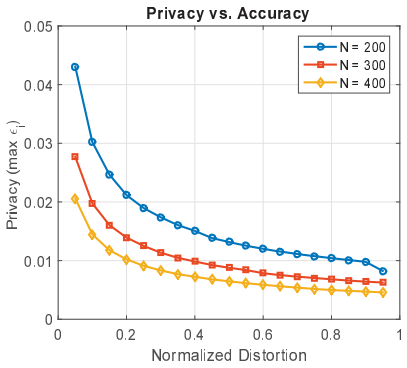

Figure 3: Relationship between data privacy and the accuracy.

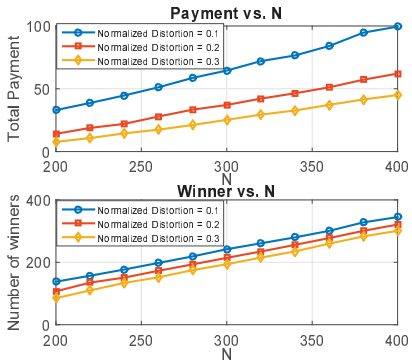

Figure 4: Effect of externalities.

only the performance of the DPDA algorithm analyzed in Section 3.

4.2 Results and Discussions

Payment versus Accuracy. In Fig. 2, we illustrate the payments under different accuracy requirements with different total number of workers. We observe that as the distortion level increases, the total payments decrease, simply because W decreases as Δ increases, i.e., the platform does not need to purchase much privacy from workers. Meanwhile, for the same level of distortion, the total payments increase with the number of workers, because W increases with the number of workers for the same level of distortion based on (13), which requires the platform to select more workers and thereby the total payments increase.

Privacy versus Accuracy. In Fig. 3, we illustrate the relationship between the data privacy and the accuracy. As the privacy of each worker is different, we use the maximum of all the workers' ϵ_i ($\epsilon = \max_{i \in \mathcal{S}} \epsilon_i$) to denote the privacy protection level at the given distortion level. As expected, as the distortion level increases, the privacy protection level increases (the smaller ϵ, the higher the privacy protection level), which agrees with our analysis in Section 2.4.

Externalities. In Fig. 4, we illustrate the effect of externalities. As discussed in Section 3.1, the data privacy of each worker depends on other workers' participations, and when the number of workers changes, it would change the privacy of each worker. As the number of workers increases, the platform needs to hire more workers to maintain the same distortion level. Therefore, we observe the increase of total payments and the number of winners. Moreover, the higher the distortion level, the lower the total payment and the less the number of winners.

Approximation. In Table 1, we illustrate the performance of the proposed DPDA algorithm by comparing the total payment by DPDA with the optimal one. For each N, we run 100 experiments and in each experiment, we randomly generate the parameters as mentioned in Section 4.1. Under different settings, we observe that the total payments generated by the DPDA algorithm is very close to the optimal one and the maximal approximation ratio for each case is around 2.

Table 1: Approximation ratio of the DPDA algorithm, where we choose the normalized distortion equal to 0.2.

Number of workers N	200	300	400
Average approximation ratio	1.88	1.85	1.85
Minimal approximation ratio	1.71	1.69	1.70
Maximal approximation ratio	2.15	2.07	1.99

Truthfulness. In Fig. 5, we verify the truthfulness of the proposed DPDA algorithm. We randomly select a winner and a loser in the auction. We fix the bids of the other workers and manipulate the selected worker's bid to evaluate the utility. Fig. 5 illustrates how the utility of the selected worker changes with her bid. As we can see that no matter how the bid changes, a winner or a loser cannot improve her utility and that the best bidding strategy for a worker is to bid truthfully.

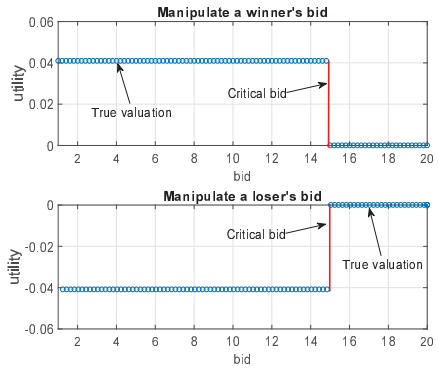

Figure 5: Truthfulness of the DPDA algorithm.

Computational Complexity. In Fig. 6, we illustrate the computational complexity of the proposed DPDA algorithm. For each N, we examine the average running time of the algorithm by running 100 experiments, in which the parameters are randomly generated as mentioned in Section 4.1. These experiments are run on a PC with a 2.7 GHz Intel Core i7 processor and 16 GB RAM. Under different settings (i.e., different distortion levels, bids, and weights), we observe that the computation time of the proposed DPDA algorithm is low and approximately linear with the problem size.

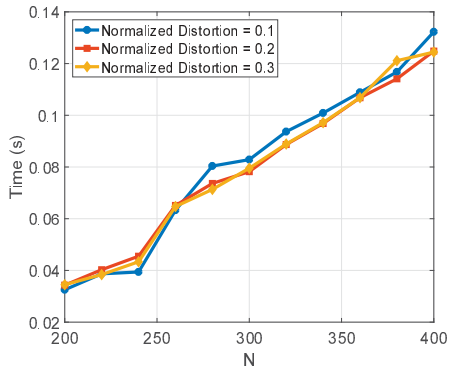

Figure 6: Computational time of the DPDA algorithm under different settings.

5 CONCLUSION AND FUTURE WORK

We studied privacy-preserving data aggregation for mobile crowdsensing in an auction framework, where the platform plays the role as an auctioneer to recruit workers to complete a sensing task. Under this model, we designed a novel mobile crowdsensing system by leveraging the concept of differential privacy. Specifically, we designed a data aggregation that allows each worker to report a noisy data and can guarantee the use of each worker's data in a differentially private manner. Then, we designed a truthful, individual rational and computationally efficient incentive mechanism that can find a set of workers to approximately minimize the cost of purchasing the private sensing data from workers subject to the accuracy requirement of the aggregated result. We validated the proposed scheme through theoretical analysis as well as extensive simulations.

We caution that although this work successfully tackled the issue of untrustworthy platform by allowing workers to conduct data perturbation locally, the workers still do not have full control of their data privacy. In this model, each worker is willing to participate as long as the individual rationality requirement is satisfied (her privacy loss got compensated). The noise parameter for each worker is determined by the externalities induced from the workers' participation, thereby indirectly determined by the platform. In practice, workers may refuse to participate if they are allowed to add only a low level of noise into data, even if they can get large enough monetary compensation in return.

In future work, we will tackle this issue by resorting to an alternative model that grants workers more power to decide their privacy-preserving levels. A possible choice could be an auction model, where besides bidding the unit privacy cost, potential workers would bid their desired privacy levels, based on which the platform selects out the set of workers for the sensing task. And we will need to evaluate workers' strategic behaviors of reporting their privacy levels and guarantee the truthfulness of the designed mechanism.

6 ACKNOWLEDGMENTS

This work is supported in part by the U.S. National Science Foundation under Grants CNS-1559696, IIA-1301726, CNS-1566634, ECCS-1711991, ECCS-1408409 and SaTC-1618768, and in part by NSFC under Grant 61731004 and Open Project of State Key Laboratory of Industrial Control Technology under Grant ICT1800373.

APPENDIX: PROOF OF THEOREM 4

To show Theorem 4, we will first characterize the optimal allocation $\mathbf{x}^{R*}$ of problem (14), based on which we can establish the relationship between the outputs of DPDA and the optimal solution to problem (12).

First, we need to establish the equivalence between problem (14) and the following problem:

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} b_i w_i x_i - C \sum_{i=1}^{N} w_i(1-x_i) \\ \text{subject to} & \sum_{i\in\mathcal{N}} w_i x_i \geq W, \\ & 0 \leq x_i \leq 1, \forall i \in \mathcal{N}. \end{array} \tag{16}$$

Lemma 3. *The optimal allocation $\mathbf{x}^{R*}$ of problem (14) is the same as the optimal allocation of problem (16).*

Proof. Since problem (14) and problem (16) have the same feasible set, any feasible $\mathbf{x}$ for problem (14) must be feasible for (16). Note that C is the optimal value of problem (14), i.e., $C = \sum_{i=1}^{N} b_i w_i x_i^{R*} / \sum_{i=1}^{N} w_i(1-x_i^{R*})$. Therefore, for any feasible $\mathbf{x}$, we have $C \leq \sum_{i=1}^{N} b_i w_i x_i / \sum_{i=1}^{N} w_i(1-x_i)$. By some algebra, we have $\sum_{i=1}^{N} b_i w_i x_i - C \sum_{i=1}^{N} w_i(1-x_i) \geq 0$, which holds with equality at $\mathbf{x}^{R*}$. In other words, $\mathbf{x}^{R*}$ is the optimal solution to problem (16). Therefore, the lemma holds. □

Based on Lemma 3, to characterize the optimal allocation $\mathbf{x}^{R*}$, we can characterize the optimal solution to problem (16) by leveraging the monotonicity property, which is given by the following lemma.

Lemma 4. *Define $l = \max\{j : \sum_{i=1}^{j} b_i w_i - C(\sum_{i=j+1}^{N} w_i) \leq 0, \forall j = 1, \ldots, N\}$. The optimal solution $\mathbf{x}^{R*}$ to problem (16) is given as follows:*

$$x_i^{R*} = \begin{cases} 1, & if\ i \leq l \\ \frac{C\sum_{i=l+1}^{N} w_i - \sum_{i=1}^{l} b_i w_i}{(b_{l+1}+C)w_{l+1}}, & if\ i = l+1 \\ 0, & if\ i > l+1 \end{cases}. \tag{17}$$

Proof. First, we show the existence of l and x_i^{R*}. Define $p(k) = \sum_{i=1}^{k} b_i w_i - C(\sum_{i=k+1}^{N} w_i)$, where $k = 0, 1, \ldots, N$. If $k = 0$, $p(0) = -C\sum_{i=1}^{N} w_i < 0$; if $k = N$, $p(N) = \sum_{i=1}^{N} b_i w_i > 0$. Note that $p(k)$ is strictly increasing with k. Therefore, there exists l such that $p(l) \leq 0$ and $p(l+1) > 0$. Further, define $q(x_{l+1}) = p(l) + (b_{l+1}w_{l+1} + Cw_{l+1})x_{l+1}$ with $0 \leq x_{l+1} \leq 1$, where $q(0) = p(l) \leq 0$ and $q(1) = p(l+1) > 0$. As $q(x_{l+1})$ is continuous and strictly increasing with x_{l+1}, there exists a unique $x_{l+1}^{R*} \in [0,1]$ such that $q(x_{l+1}^{R*}) = 0$. Next, to verify the optimal solution $\mathbf{x}^{R*}$ in (17), we can leverage the KKT conditions for problem (16). Specifically, the Lagrangian for problem (16) is

$$\begin{array}{l} L(\mathbf{x}, \lambda, \mu, \nu) = \sum_{i=1}^{N} b_i w_i x_i - C\sum_{i=1}^{N} w_i(1-x_i) \\ +\lambda(W - \sum_{i=1}^{N} w_i x_i) + \sum_{i=1}^{N} \mu_i(x_i - 1) - \sum_{i=1}^{N} \nu_i x_i \\ = \sum_{i=1}^{N}(b_i w_i + Cw_i - \lambda w_i + \mu_i - \nu_i)x_i \\ -C\sum_{i=1}^{N} w_i + \lambda W - \sum_{i=1}^{N} \mu_i, \end{array} \tag{18}$$

where λ, μ and ν are Lagrangian multipliers. From (18), we have $b_i w_i + Cw_i - \lambda w_i + \mu_i - \nu_i = 0$ for all x_i. It is easy to verify that x_i^{R*} given by Lemma 4 satisfies KKT conditions with $\lambda^* = b_{l+1} + C$,

$\mu_i^* = 1_{\{i \le l\}}(b_{l+1} - b_i)w_i$, and $v_i^* = 1_{\{i > l+1\}}(b_i - b_{l+1}w_i)$, where $1_{condition}$ denotes the indicator function, i.e., $1_{condition} = 1$ if the condition holds and $1_{condition} = 0$, otherwise. Therefore, the lemma holds. □

Based on Lemma 4, it is easy to show that DPDA chooses $l+1$ workers and satisfies the accuracy requirement.

LEMMA 5. *Let l be as defined in Lemma 4 and k be the number of winners chosen by DPDA. Then, we have $k = l+1$ and DPDA satisfies the accuracy requirement (i.e., $\delta(\mathbf{x}) \le \Delta$).*

PROOF. By construction, DPDA will choose the smallest k such that $\sum_{i=1}^{k} b_i w_i - C(\sum_{i=k+1}^{N} w_i) \ge 0$, which are workers $1, 2, \ldots, l+1$ from Lemma 4. Moreover, $\sum_{i=1}^{l+1} w_i \ge \sum_{i=1}^{l} w_i + w_{l+1}x_{l+1}^{R*} \ge W$, since $x_{l+1}^{R*} \in [0,1]$, which concludes the proof. □

Now, we will show that DPDA is α-approximation with respect to the optimal total payment, which is equivalent to showing that DPDA is α-approximation with respect to the optimal value of problem (16), based on Lemma 1 and Lemma 3. Note that the objective function in (16) contains a constant $C\sum_{i=1}^{N} w_i$ and removing this constant will not impact the result of the optimization problem (16). In other words, we can focus on the approximation of DPDA with respect to the function $h(\mathbf{x}) = \sum_{i=1}^{N}(b_i + C)w_i x_i$. Let $\mathbf{x}^{DPDA}$ be the allocation by DPDA and $\mathbf{x}^*$ be the optimal allocation with OPT $= h(\mathbf{x}^*)$ Since problem (16) is a relaxation of problem (13), $h(\mathbf{x}^{R*}) \le$ OPT. In what follows, we will show $h(\mathbf{x}^{DPDA}) \le \alpha$OPT.

Based on Lemma 5, we have

$$\begin{aligned} h(\mathbf{x}^{DPDA}) &= \sum_{i=1}^{l+1}(b_i + C)w_i \\ &= \sum_{i=1}^{l}(b_i + C)w_i + (b_{l+1} + C)w_{l+1}x_{l+1}^{R*} \\ &\quad + (b_{l+1} + C)w_{l+1}(1 - x_{l+1}^{R*}) \\ &= h(\mathbf{x}^{R*}) + (b_{l+1} + C)w_{l+1}(1 - x_{l+1}^{R*}), \end{aligned} \tag{19}$$

where $h(\mathbf{x}^{R*}) = \sum_{i=1}^{l}(b_i + C)w_i + (b_{l+1} + C)w_{l+1}x_{l+1}^{R*}$ based on (17). Then,

$$\begin{aligned} \frac{h(\mathbf{x}^{R*}) + (b_{l+1}+C)w_{l+1}(1-x_{l+1}^{R*})}{h(\mathbf{x}^{R*})} &= 1 + \frac{(b_{l+1}+C)w_{l+1}(1-x_{l+1}^{R*})}{h(\mathbf{x}^{R*})} \\ &\overset{(a)}{\le} 1 + \frac{(b_{l+1}+C)w_{l+1}(1-x_{l+1}^{R*})}{(b_{l+1}+C)w_{l+1}x_{l+1}^{R*}} \\ &= \frac{1}{x_{l+1}^{R*}} = \alpha, \end{aligned} \tag{20}$$

where (a) follows from the fact that $h(\mathbf{x}^{R*}) \ge (b_{l+1} + C)w_{l+1}x_{l+1}^{R*}$. Since $x_{l+1}^{R*} \le 1$, we have $\alpha \ge 1$. Therefore, $h(\mathbf{x}^{DPDA}) \le \alpha h(\mathbf{x}^{R*}) \le \alpha$OPT, which concludes the proof.

REFERENCES

[1] CHEUNG, M. H., SOUTHWELL, R., HOU, F., AND HUANG, J. Distributed time-sensitive task selection in mobile crowdsensing. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (2015), ACM, pp. 157–166.

[2] CHRISTIN, D., REINHARDT, A., KANHERE, S. S., AND HOLLICK, M. A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software 84*, 11 (2011), 1928–1946.

[3] DANDEKAR, P., FAWAZ, N., AND IOANNIDIS, S. Privacy auctions for recommender systems. *ACM Transactions on Economics and Computation 2*, 3 (2014), 12.

[4] DUAN, L., KUBO, T., SUGIYAMA, K., HUANG, J., HASEGAWA, T., AND WALRAND, J. Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing. In *INFOCOM, 2012 Proceedings IEEE* (2012), IEEE, pp. 1701–1709.

[5] DUAN, L., KUBO, T., SUGIYAMA, K., HUANG, J., HASEGAWA, T., AND WALRAND, J. C. Motivating smartphone collaboration in data acquisition and distributed computing. *IEEE Trans. Mob. Comput. 13*, 10 (2014), 2320–2333.

[6] DWORK, C. Differential privacy. In *Automata, Languages and Programming*, vol. 4052 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2006, pp. 1–12.

[7] FENG, Z., ZHU, Y., ZHANG, Q., NI, L. M., AND VASILAKOS, A. V. TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In *INFOCOM, 2014 Proceedings IEEE* (2014), IEEE, pp. 1231–1239.

[8] GAO, L., HOU, F., AND HUANG, J. Providing long-term participation incentive in participatory sensing. In *Computer Communications (INFOCOM), 2015 IEEE Conference on* (2015), IEEE, pp. 2803–2811.

[9] GHOSH, A., AND ROTH, A. Selling privacy at auction. *Games and Economic Behavior 91* (2015), 334–346.

[10] GONG, X., AND SHROFF, N. B. Truthful mobile crowdsensing for strategic users with private qualities. In *WiOpt 2017, Paris, France, May 15-19* (2017).

[11] HE, S., SHIN, D.-H., ZHANG, J., AND CHEN, J. Toward optimal allocation of location dependent tasks in crowdsensing. In *Proc. IEEE INFOCOM* (2014).

[12] HE, S., SHIN, D.-H., ZHANG, J., JIMING, C., AND LIN, P. An exchange market approach to mobile crowdsensing: Pricing, task allocation and walrasian equilibrium. *IEEE Journal on Selected Areas in Communications* (2017).

[13] IBM ILOG. Introducing IBM ILOG CPLEX optimization studio v12.5.1. http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r5/index.jsp.

[14] JIN, H., SU, L., CHEN, D., NAHRSTEDT, K., AND XU, J. Quality of information aware incentive mechanisms for mobile crowd sensing systems. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (2015), ACM, pp. 167–176.

[15] JIN, H., SU, L., XIAO, H., AND NAHRSTEDT, K. Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems. In *Proceedings of the 17th international symposium on mobile Ad Hoc networking and computing, MobiHoc* (2016), vol. 16, pp. 341–350.

[16] KOTZ, S., KOZUBOWSKI, T., AND PODGORSKI, K. *The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance*. Springer Science & Business Media, 2012.

[17] LI, Q., LI, Y., GAO, J., ZHAO, B., FAN, W., AND HAN, J. Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data* (2014), ACM, pp. 1187–1198.

[18] MEGIDDO, N. *On the complexity of linear programming*. IBM Thomas J. Watson Research Division, 1986.

[19] MENG, C., JIANG, W., LI, Y., GAO, J., SU, L., DING, H., AND CHENG, Y. Truth discovery on crowd sensing of correlated entities. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems* (2015), ACM, pp. 169–182.

[20] MILGROM, P. R. *Putting auction theory to work*. Cambridge University Press, 2004.

[21] PENG, D., WU, F., AND CHEN, G. Pay as how well you do: A quality based incentive mechanism for crowdsensing. In *Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing* (2015), ACM, pp. 177–186.

[22] SHENG, X., TANG, J., XIAO, X., AND XUE, G. Sensing as a service: Challenges, solutions and future directions. *IEEE Sensors journal 13*, 10 (2013), 3733–3741.

[23] WANG, W., YING, L., AND ZHANG, J. Buying data from privacy-aware individuals: the effect of negative payments. In *International Conference on Web and Internet Economics* (2016), Springer, pp. 87–101.

[24] WANG, W., YING, L., AND ZHANG, J. The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits. In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science* (2016), ACM, pp. 249–260.

[25] WEN, Y., SHI, J., ZHANG, Q., TIAN, X., HUANG, Z., YU, H., CHENG, Y., AND SHEN, X. Quality-driven auction-based incentive mechanism for mobile crowd sensing. *IEEE Transactions on Vehicular Technology 64*, 9 (2015), 4203–4214.

[26] YANG, D., XUE, G., FANG, X., AND TANG, J. Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing. In *Proceedings of the 18th annual international conference on Mobile computing and networking* (2012), ACM, pp. 173–184.

[27] ZHANG, M., YANG, L., GONG, X., AND ZHANG, J. Privacy-preserving crowdsensing: Privacy valuation, network effect, and profit maximization. In *Global Communications Conference (GLOBECOM), 2016 IEEE* (2016), IEEE, pp. 1–6.

[28] ZHANG, Q., WEN, Y., TIAN, X., GAN, X., AND WANG, X. Incentivize crowd labeling under budget constraint. In *Computer Communications (INFOCOM), 2015 IEEE Conference on* (2015), IEEE, pp. 2812–2820.

[29] ZHANG, X., XUE, G., YU, R., YANG, D., AND TANG, J. Truthful incentive mechanisms for crowdsourcing. In *Computer Communications (INFOCOM), 2015 IEEE Conference on* (2015), IEEE, pp. 2830–2838.

[30] ZHAO, D., LI, X.-Y., AND MA, H. How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint. In *INFOCOM, 2014 Proceedings IEEE* (2014), IEEE, pp. 1213–1221.