DPDT: A Differentially Private Crowd-Sensed Data Trading Mechanism

Guoju Gao, Mingjun Xiao, *Member, IEEE*, Jie Wu, *Fellow, IEEE*, Sheng Zhang, *Member, IEEE*, Liusheng Huang, *Member, IEEE*, and Guiliang Xiao

***Abstract*—Along with the generation of Internet of Things (IoT), the values of tremendous volumes of sensing data will be slowly unlocked. Thus, crowd-sensed data trading as a new business paradigm has recently attracted increasing attention. A typical data trading system contains a platform, data consumers, and crowd workers. The platform recruits crowd workers to collect data and then sells the data to consumers. In this article, we design a differentially private crowd-sensed data trading mechanism, called DPDT, to preserve the identity privacy of consumers and the task privacy against crowd workers during the data collection process, simultaneously. DPDT consists of a differentially private auction-based data pricing algorithm and a differentially private data collection algorithm. The data pricing algorithm achieves a good approximation to the maximum revenue. Meanwhile, it guarantees $(e^2-1)\epsilon$-truthfulness and 2ϵ-differential privacy, where $\epsilon > 0$ is a small constant. The data collection algorithm is able to effectively protect the data collection task privacy against crowd workers. We prove that this data collection algorithm achieves δ-approximate ϵ-differential privacy, where $\delta < 1/e$ is a small constant, and meanwhile guarantees a tight bound of the expected approximation ratio. At last, extensive simulations are conducted to verify the significant performance of DPDT.**

***Index Terms*—Approximate truthfulness, auction, data trading, differential privacy, mobile crowdsensing.**

I. Introduction

RECENTLY, with the proliferation of mobile devices equipped with more and more components, a new sensing paradigm called mobile crowdsensing has been proposed [6], [33]. Since, mobile crowdsensing can coordinate a group of mobile users to collect tremendous volumes of sensed data (such as traffic condition monitoring, noise pollution monitoring, wireless indoor location, etc.) over an urban environment that individual users cannot cope with, it has attracted much attention. Essentially speaking, mobile crowdsensing is a special form of data trading [7], [15], [23], [39] where mobile users get some monetary reward by sharing their collected data. Generally, in the data trading market, data consumers can access sufficient data to conduct some research, while data providers will obtain some monetary reward. Therefore, data trading has huge commercial value and bright application prospects. To fully deliver the potential of the data trading market, more and more data trading platforms (such as DataExchange, Datacoup, Terbine, CitizenMe, Thingspeak, etc.) have emerged to enable crowd-sensed data to be exchanged on the Web.

A typical data trading system mainly consists of three parts: 1) a platform; 2) data consumers; and 3) crowd workers (also known as providers), as shown in Fig. 1. The platform recruits crowd workers to collect sensed data and then sells the data to consumers. The ultimate goal of the data trading system is to maximize its profit, i.e., the difference between the revenue of selling data to consumers and the cost of recruiting crowd workers. So far, there has been some research on the crowd-sensed data trading problem [10], [15], [39]. For example, He *et al.* [10] studied an exchange market approach to mobile crowdsensing; Jung *et al.* [15] proposed some accountable protocols for big data trading against dishonest consumers; Zheng *et al.* [39] designed the profit-driven data acquisition scheme for crowd-sensed data market. Nevertheless, these existing works rarely involve the significant privacy-preserving issues.

In this article, we focus on designing a privacy-preserving crowd-sensed data trading mechanism to maximize the profits of the platform, where the identity privacy of the data consumers (i.e., buyers) and the task privacy against crowd workers during the data collection process can be protected effectively. In fact, only a few literatures [7], [23], [37] show concern for privacy issues in the crowd-sensed data trading market. Among them, [37] protects image privacy based on the concept of feature-indistinguishability (i.e., MinHash mechanism), while [7] and [23] adopt homomorphic encryption (and partial electronic signature technique) to protect either the bid privacy or the identity privacy of data contributors (i.e., sellers). Here, the homomorphic encryption requires high computation complexity, and the MinHash mechanism is actually used to generate an electronic signature. Thus, both of

Manuscript received August 8, 2019; revised August 19, 2019 and September 16, 2019; accepted September 24, 2019. Date of publication September 26, 2019; date of current version January 10, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61872330, Grant 61572457, Grant 61379132, Grant U1709217, Grant 61303206, Grant 61572342, and Grant 61502261, in part by NSF under Grant CNS 1824440, Grant CNS 1828363, Grant CNS 1757533, Grant CNS 1618398, Grant CNS 1651947, and Grant CNS 1564128, in part by the NSF of Jiangsu Province in China under Grant BK20191194, Grant BK20131174, and Grant BK2009150, and in part by the Anhui Initiative in Quantum Information Technologies under Grant AHY150300. *(Corresponding author: Mingjun Xiao.)*

G. Gao, M. Xiao, L. Huang, and G. Xiao are with the School of Computer Science and Technology/Suzhou Institute for Advanced Study, University of Science and Technology of China, Hefei 230026, China (e-mail: xiaomj@ustc.edu.cn).

S. Zhang is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China (e-mail: sheng@nju.edu.cn).

J. Wu is with the Department of Computer and Information Sciences, Temple University, Philadelphia, PA 19122 USA (e-mail: jiewu@temple.edu).

Digital Object Identifier 10.1109/JIOT.2019.2944107

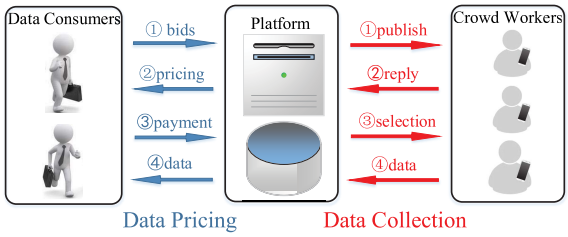


Fig. 1. Illustration of the data trading scenario.

them do not apply to our crowd-sensed data trading scenario. Moreover, our data trading needs to consider both the identity privacy (data pricing) and task privacy (data collection). Furthermore, data pricing also involves competition among data consumers. In this article, we study how to design an efficient privacy-preserving crowd-sensed data trading mechanism. In fact, designing such a mechanism has the following challenges.

First, due to the competition among data consumers, as well as the selfishness and individual rationality of consumers, a truthful (also known as strategy-proof) auction mechanism which motivates bidders to report their true valuation as bids [5], [31], [34] is indispensable here. However, the absolute truthfulness may lower the revenues of the platform [22], [42], so we need to properly relax the notion of absolute truthfulness. Finding the tradeoff between revenue maximization and truthfulness is challenging.

Second, reporting the true valuations (i.e., the submitted bids) on the data will reveal consumers' identity-related information, which might result in the disclosure of the critical commercial secrets [7], [8], [26], [39]. One (other than the platform) may infer the private valuation of a bidder according to the outcomes of auction. In particular, the data (e.g., traffic monitoring data in a certain location) is constantly updated over time. Therefore, the data consumers would soon compete again for the updated data of the same kind. This makes the derivation of the bidders' identities easier [4], [42].

Third, the crowd-sensed data to be sold may reveal the ongoing work of consumers, which will also incur a privacy leakage problem [21], [24], [35]. Since the crowd-sensed data in the platform is transparent to consumers, data privacy is not a concern among consumers. However, data privacy may be leaked to crowd workers during the collection process. Unfortunately, it is fairly difficult to minimize the recruitment cost while guaranteeing the privacy of the collected data.

In this article, to address the above challenges, we propose a differentially private crowd-sensed data trading mechanism, called DPDT. DPDT consists of a differentially private auction-based data pricing algorithm and a differentially private data collection algorithm. First, in order to balance the revenue maximization and truthfulness [22], [42] and at the same time protect the identity privacy of data consumers, we combine the concept of approximate truthfulness with the differential privacy (an exponential mechanism). Then, we design a differentially private auction-based data pricing algorithm for DPDT, which achieves a good approximation to the maximum revenue, preserves the identity privacy of consumers, and also guarantees that data consumers have limited incentives to lie. Second, in order to minimize the cost of collecting data while protecting the data collection task privacy against crowd workers, we model the data collection problem as a special set cover problem with differential privacy. Based on this, we devise a differentially private data collection algorithm for DPDT. To the best of our knowledge, we are the first to study the consumers' identity privacy, truthfulness, and the data collection task privacy in the data trading field. Our major contributions are summarized as follows.

1) We propose a differentially private crowd-sensed data trading mechanism, consisting of a differentially private auction-based data pricing algorithm and a differentially private data collection algorithm. Both the consumers' identity privacy and the data collection task privacy can be protected effectively by DPDT.
2) The data pricing algorithm not only achieves 2ϵ-differential privacy, but also guarantees $(e^2-1)\epsilon$-truthfulness, where $\epsilon > 0$ is a small constant. Additionally, this data pricing algorithm can achieve an expected revenue of at least $\text{opt} - 3\ln(e+\epsilon|\mathbb{P}|\text{opt})/\epsilon$, where opt is the optimal revenue and $\mathbb{P}$ is the set of possible prices.
3) The data collection algorithm can obtain δ-approximate ϵ-differential privacy where $\delta < 1/e$ is a small constant, and at the same time can achieve an expected approximation ratio of $O(\ln|\mathbb{U}|+(\ln(|\mathbb{W}|\ln(e/\delta)))/\epsilon)$, in which $\mathbb{U}$ and $\mathbb{W}$ are the sets of total data collection tasks and crowd workers, respectively.
4) We conduct extensive simulations to evaluate the performance of DPDT. The simulation results show that DPDT can obtain good revenues, and can also effectively protect the identity privacy of consumers and the task privacy during the data collection process.

The remainder of this article is organized as follows. We first describe the crowd-sensed data trading model and introduce some related solution concepts in Section II. Then, we design the differentially private auction-based data pricing algorithm and the differentially private data collection algorithm in Sections III and IV, respectively. In Section V, we evaluate the performance of the proposed algorithms. After reviewing the related work in Section VI, we conclude this article in Section VII.

II. System Model

In this section, we first describe the overview of the crowd-sensed data trading system. Then, we present detailed data pricing and data collection modules, respectively. Finally, we introduce some relevant solution concepts about differential privacy and the auction theory.

A. Crowd-Sensed Data Trading System

We consider a typical crowd-sensed data trading system, as shown in Fig. 1, which is mainly composed of a platform in a cloud, multiple registered data consumers (e.g., government departments, companies, individuals, etc.) and lots of

registered crowd workers. The platform mainly consists of a control center and a data pool, and actually acts as a bridge between data consumers and crowd workers. The platform recruits crowd workers to collect some sensed data which will be sold to data consumers in the future. It aims to maximize the profit, which is defined as the difference between the revenue of selling data to consumers and the cost of recruiting crowd workers. Thus, the platform concentrates on simultaneously maximizing the revenue and minimizing the recruitment cost, while protecting the identity privacy of consumers and task privacy against crowd workers during the data collection process. Based on this, the data trading system can be divided into two separate modules: 1) data pricing and 2) data collection.

B. Data Pricing in DPDT

In the data pricing module, the platform focuses on setting an appropriate price for the owned data so that it can maximize its revenue, where the revenue is the total payment received from consumers. At the same time, the identity privacy preservation for data consumers and the competition among consumers should be taken into consideration. Consequently, we combine differential privacy with auction in the data pricing module. We consider a collusion-free auction where the platform also acts as the auctioneer. Note that the platform might hold multiple datasets to be sold, and each dataset will be set an independent price. For simplicity of the following descriptions, we only consider one dataset in the platform, and we let the platform conduct the same operations for other datasets in the practical scenario.

Consider that a set of data consumers registered in the platform, denoted by $\mathbb{N} = \{1, 2, \ldots, n\}$, will compete for the dataset. Each consumer (also known as buyer or bidder) $i \in \mathbb{N}$ has a valuation on the dataset, denoted by v_i. He also determines a bid b_i (the claimed valuation), and sends b_i to the platform. The platform only knows the claimed valuation b_i instead of the true valuation v_i. In fact, v_i is known to nobody except bidder i itself. Bidder i may strategically manipulate v_i to get a higher utility. Such strategic manipulation might reduce the revenues of the platform. Thus, the data pricing algorithm must ensure that each bidder will not manipulate its bids, i.e., *truthfulness*. However, absolute truthfulness may also lower the revenues of the platform [22], [42]. Accordingly, to maximize the revenues while ensuring that bidders have limited incentives to lie, a concept of *approximate truthfulness* [22], [42] is adopted here. The detailed definition will be described in Section II-D.

For simplicity of the following computation, all values of v_i and b_i for $i \in \mathbb{N}$ are normalized into $(0, 1]$. Let $\mathbb{V} = \{v_1, v_2, \ldots, v_n\}$ and $\mathbb{B} = \{b_1, b_2, \ldots, b_n\}$ denote the valuation and bid sets, respectively. Note that the claimed valuation b_i is closely related to the sensitive identity information, which needs to be kept private [7], [39]. On the other hand, the auctioneer (i.e., platform) initializes a finite set of candidate prices $\mathbb{P} = \{p_1, p_2, \ldots, p_\kappa\}$, including all the possible valuation/bid values in $(0, 1]$.

After receiving the bid set $\mathbb{B}$, the platform selects a feasible price $p \in \mathbb{P}$, determines the winning bid set, and further computes the payment for winning bidders. Each bidder $i \in \mathbb{N}$, as a selfish and rational person, will always maximize its utility. In fact, a bidder's (e.g., i) utility depends on several factors: his bid b_i and valuation v_i, other bidders' bids $\mathbb{B}_{-i}$ (here, $\mathbb{B}_{-i}$ means the bid set except i, i.e., $\mathbb{B} = \mathbb{B}_{-i} + \{b_i\}$), and the selected price p. Note that the winning bid set depends on b_i and $\mathbb{B}_{-i}$, while the payment for winning bids relies on the value p. Here, we let $u_i(\cdot)$ denote the utility of bidder i, which is calculated as follows:

$$u_i(b_i, \mathbb{B}_{-i}, v_i, p) = \begin{cases} 0; & b_i \le p \\ v_i - p; & b_i > p. \end{cases} \tag{1}$$

Problem Formulation: The platform concentrates on maximizing its revenues, i.e., the total payment from buyers, under the privacy restrictions. Concretely speaking, the platform aims at selecting one of the prices according to the received bids $\mathbb{B}$, so that it can maximize its revenues, while protecting the identity (i.e., bid) privacy of bidders and guaranteeing the approximate truthfulness, simultaneously.

C. Data Collection in DPDT

In the data collection module, the platform concentrates on selecting a minimum number of crowd workers to collect sensed data. The fewer the number of workers, the less the recruitment cost. Since the collected data will be sold to consumers in the future, the sensed data should be kept private from crowd workers during the collection process. To this end, we need to protect the data collection tasks from being directly revealed to crowd workers. We first adopt the perturbation method in the data collection task publishing process and then apply the differential privacy (an exponential mechanism) in the worker recruitment process.

In the task publishing process, consider that the platform has a total of R types of crowd-sensed data collection tasks. These tasks are distributed at different locations. We use $\mathbb{R}$ to denote the task set. In order to protect the data privacy against crowd workers, the platform cannot publish $\mathbb{R}$ directly. For this reason, we first generate a set of noisy tasks (denoted by $\mathbb{T}$), and then add $\mathbb{T}$ into $\mathbb{R}$. Instead of $\mathbb{R}$, the platform will publish $\mathbb{R} \cup \mathbb{T}$ to the crowd workers to hide its true task set $\mathbb{R}$ (called perturbation). Generally, the distributions of $\mathbb{R}$ and $\mathbb{T}$ are similar, and the cardinalities of $\mathbb{R}$ and $\mathbb{T}$ are on the same order of magnitude. For simplicity, we use $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$ to denote the total task set, and let $\mathbb{W}$ denote the set of crowd workers. Then, each worker $k \in \mathbb{W}$ replies to the platform with the set of tasks that it can complete, denoted by $\mathbb{G}_k \subseteq \mathbb{U}$. We use $\Omega = \{\mathbb{G}_k | k \in \mathbb{W}\}$ to denote the returned results of all workers. It is inevitable that the total task set $\mathbb{U}$ after adding noisy tasks will burden crowd workers more. We assume that the data consumers will bear these additional expenses, since the fundamental reason for the platform generating noisy tasks is to protect the data privacy of consumers. Actually, the size of the noisy task set $\mathbb{T}$ will have an effect on the achieved differential privacy level and the expected approximation ratio, simultaneously. We will analyze the impact of $\mathbb{T}$ in detail in Section IV.

Problem Formulation: In the worker recruitment process, based on the true task set $\mathbb{R}$, the total task set $\mathbb{U}$, and the collection $\Omega = \{\mathbb{G}_k | k \in \mathbb{W}\}$, the platform focuses on selecting a minimum number of workers (denoted by $\Phi \subseteq \mathbb{W}$) so that $\mathbb{R}$ can be *covered*, while $\mathbb{R}$ is kept private against all crowd workers. "Covered" here means that the data collection tasks in $\mathbb{R}$ can be performed by crowd workers in Φ successfully.

D. Solution Concepts

In this section, we introduce some related solution concepts from differential privacy and auction theory. First, we present the definition of differential privacy.

Definition 1 (Differential Privacy) [4], [9], [25]: A randomized mechanism $\mathcal{M}$ has ϵ-differential privacy if for two input profiles D_1 and D_2 where the two input profiles differ in a single element, and for all outcomes $M \subseteq \text{Range}(\mathcal{M})$

$$\Pr[\mathcal{M}(D_1) \in M] \leq \exp(\epsilon) \times \Pr[\mathcal{M}(D_2) \in M] \tag{2}$$

where $\epsilon > 0$ is the privacy budge/level (a small constant).

Moreover, we also introduce a laxation of differential privacy, which allows a small additive term in the bound.

Definition 2 (Approximate Differential Privacy) [4], [9]: We say a randomized mechanism $\mathcal{M}$ has δ-approximate ϵ-differential privacy if for two input profiles D_1 and D_2 differing in one element, and for all outcomes $M \subseteq \text{Range}(\mathcal{M})$

$$\Pr[\mathcal{M}(D_1) \in M] \leq \exp(\epsilon) \times \Pr[\mathcal{M}(D_2) \in M] + \delta \tag{3}$$

where $\delta > 0$ is a small constant.

Second, we present a powerful technique used in differential privacy, called exponential mechanism [11], [22], which can construct differentially private algorithms over an arbitrary range $\mathbb{P}$ of outcomes and any object function $Q(\mathbb{B}, p)$ that maps a pair consisting of a bid set $\mathbb{B}$ and a feasible outcome $p \in \mathbb{P}$ to a real-valued score. Note that here $\mathbb{B}$ is the claimed valuation set of bidders, and p is the payment for each winning bidder. Thus, we have the following definition.

Definition 3 (Exponential Mechanism) [11], [22]: Given a range $\mathbb{P}$, a bid set $\mathbb{B}$, a revenue function Q, and a privacy parameter ϵ, the exponential mechanism $\text{Exp}(\mathbb{P}, \mathbb{B}, Q, \epsilon)$ selects an outcome p from $\mathbb{P}$ with the probability

$$\Pr\big[\text{Exp}(\mathbb{P}, \mathbb{B}, Q, \epsilon) = p\big] \propto \exp\Big(\frac{\epsilon}{2\Delta} Q(\mathbb{B}, p)\Big) \tag{4}$$

where Δ is the Lipschitz constant of the revenue function Q, i.e., for any two adjacent input data profiles $\mathbb{B}_1$ and $\mathbb{B}_2$, and for any outcome p in the range $\mathbb{P}$, the scores $Q(\mathbb{B}_1, p)$ and $Q(\mathbb{B}_2, p)$ differ by at most Δ.

Third, we introduce the concept of approximate truthfulness in auction theory. Before defining the approximate truthfulness, we first review the dominant strategy (i.e., truthfulness) in an auction mechanism [34], [42]. That is, the strategy s_i of bidder i is a dominant strategy, if for any strategy $s_i' \neq_i$ and other bidders' strategy profile s_{-i}, then $u_i(s_i, s_{-i}) \geq_i (s_i', s_{-i})$. Based on this, we have the following definition.

Definition 4 (γ-Truthful) [42]: The expected utility of bidder i based on the dominant strategy s_i is denoted by $E[u_i(s_i, s_{-i})]$. Then, for any strategy $s_i' \neq_i$ and other bidders' strategy profile s_{-i}, we say that a mechanism is γ-truthful, if we have

$$E\big[u_i(s_i, s_{-i})\big] \geq E\big[u_i\big(s_i', s_{-i}\big)\big] - \gamma \tag{5}$$

in which $\gamma > 0$ is a small constant. The approximate truthfulness of the auction mechanism can ensure that each data consumer has limited incentives to lie, which may increase the total revenues of the platform.

Additionally, we summarize the commonly used notations throughout this article in Table I.

TABLE I
DESCRIPTION OF MAJOR NOTATIONS

Variable	Description
$\mathbb{N}$, $\mathbb{W}$	the sets of data consumers and crowd workers.
i, k	the indexes for consumers and workers.
v_i, b_i	the valuation and bid of the i-th data consumer.
$\mathbb{V}$, $\mathbb{B}$	the valuation and bid sets, respectively.
$\mathbb{P}$, p	the set of candidate prices and a price instance.
$\mathbb{B}_{-i}$	the set of bids except the consumer i.
$\widetilde{\mathbb{B}}(p)$	the set of bids where $\forall b_i \in \widetilde{\mathbb{B}}(p) \geq p$.
$\mathbb{R}$, $\mathbb{T}$	the sets of true and noisy data collection tasks.
$\mathbb{U}$	the sets of total tasks, i.e., $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$.
Ω	the collection of the subsets of $\mathbb{U}$ and $\lvert\Omega\rvert = \lvert\mathbb{W}\rvert$.
$Pr, Pr[\cdot]$	the probability vector and a probability value.
$Q(\mathbb{B}, p)$	the revenue function based on $\mathbb{B}$ and p.
ϵ, δ	the parameters of differential privacy.
$E[\cdot]$	the expected function.

III. DPDT: Data Pricing Algorithm

In this section, we design a differentially private auction-based data pricing algorithm, in which we combine the exponential mechanism with the auction mechanism to achieve both approximate revenue maximization and differential privacy. The algorithm is actually based on a posted pricing auction mode, where we select a price from the set of prices according to a designed probability distribution. The probability of selecting a price is proportional to the achieved revenues. In the following, we will present the calculation of probability distribution and price selection in detail. After that, we will describe the detailed algorithm and give performance analysis.

A. Probability Calculation and Price Selection

For each price $p_j \in \mathbb{P}$ and the bid set $\mathbb{B}$, we first remove the bidders whose claimed bids are less than p_j. Then, the original bid set $\mathbb{B}$ is changed to $\widetilde{\mathbb{B}}(p_j)$, that is,

$$\widetilde{\mathbb{B}}(p_j) = \big\{b_i | b_i \in \mathbb{B} \wedge b_i \geq p_j\big\}. \tag{6}$$

Since the tentative price for the candidate bidders in $\widetilde{\mathbb{B}}(p_j)$ is p_j, the platform's total revenues are calculated by

$$Q\big(\mathbb{B}, p_j\big) = p_j \times \big|\widetilde{\mathbb{B}}\big(p_j\big)\big| \tag{7}$$

where $|\cdot|$ denotes the cardinality of a set.

After calculating the revenues of the platform by setting all possible prices in $\mathbb{P}$, the algorithm computes the probability

Algorithm 1 Differentially Private Data Pricing Algorithm

Require: $\mathbb{N}$, $\mathbb{P}$, $\mathbb{B}$, ϵ

Ensure: a probability vector, selected price, and winner set.

Phase 1: platform publishes the dataset to $\mathbb{N}$;

Phase 2: data consumers submit their bid set $\mathbb{B}$;

Phase 3: platform calculates the probability distribution;

1: **for** $p \in \mathbb{P}$ **do**

2: update bidder set $\widetilde{\mathbb{B}}(p) = \{b_i | b_i \in \mathbb{B} \wedge b_i \geq p\}$;

3: calculate the revenue $Q(\mathbb{B}, p) = p \times |\widetilde{\mathbb{B}}(p)|$;

4: **end for**

5: **for** $p \in \mathbb{P}$ **do**

6: calculate probability $Pr[p] = \frac{\exp(\epsilon Q(\mathbb{B},p))}{\sum_{p_j \in \mathbb{P}} \exp(\epsilon Q(\mathbb{B},p_j))}$;

7: **end for**

Phase 4: platform determines the price and winner set;

8: select a price $p \in \mathbb{P}$ according to the vector Pr;

9: determine the winner set $\widetilde{\mathbb{B}}(p) = \{b_i | b_i \in \mathbb{B} \wedge b_i \geq p\}$;

distribution for each single price. More specifically, the probability of selecting $p \in \mathbb{P}$ is proportional to the corresponding revenues, i.e.,

$$\Pr[p \in \mathbb{P}] = \frac{\exp(\epsilon Q(\mathbb{B}, p))}{\sum_{p_j \in \mathbb{P}} \exp\big(\epsilon Q\big(\mathbb{B}, p_j\big)\big)}. \tag{8}$$

After obtaining the probability vector $\Pr = (\Pr[p_1], \Pr[p_2], \ldots, \Pr[p_{|\mathbb{P}|}])$, we choose a price $p \in \mathbb{P}$ as the auction payment for the single dataset with the corresponding probability. That is, we set the price for the data for sale successfully. Moreover, the candidate bidders in $\widetilde{\mathbb{B}}(p)$ whose bids are not less than p are the final winners. Each winner will be allocated to the dataset and be charged with the payment p.

B. Detailed Algorithm

The differentially private data pricing algorithm is shown in Algorithm 1. In the first two phases, after receiving the description about the data to be sold, each consumer submits her/his bid to the platform. In the third phase, for each price $p \in \mathbb{P}$, the platform first updates the bid set and then computes the corresponding revenue in steps 1–4. Next, the probability distribution is calculated in steps 5–7. In the fourth phase, the price for the data is determined according to the probability distribution in step 8. Based on the selected price, the winning bidder set is also determined in step 9. The computation overhead of Algorithm 1 is dominated by step 2, which can be denoted by $O(|\mathbb{N}| \cdot |\mathbb{P}|)$. Since both $\mathbb{N}$ and $\mathbb{P}$ are finite sets, Algorithm 1 has a polynomial-time computational complexity.

C. Performance Analysis

We first analyze the achieved differential privacy level in the following theorem.

Theorem 1: The proposed data pricing algorithm can guarantee 2ϵ-differential privacy.

Proof: Consider two bid sets, denoted by $\mathbb{B}_1$ and $\mathbb{B}_2$, which differ in only one bid (change, remove, or add). Then, the probability of selecting $p \in \mathbb{P}$ based on the two bid profiles $\mathbb{B}_1$ and $\mathbb{B}_2$ are denoted by $\Pr[M(\mathbb{B}_1) = p]$ and $\Pr[M(\mathbb{B}_2) = p]$, respectively. Thus, we have

$$\begin{aligned}
&\Pr\big[M(\mathbb{B}_1) = p\big] / \Pr\big[M(\mathbb{B}_2) = p\big] \\
&= \frac{\exp(\epsilon Q(\mathbb{B}_1, p))}{\exp(\epsilon Q(\mathbb{B}_2, p))} \times \frac{\sum_{p_j \in \mathbb{P}} \exp\big(\epsilon Q\big(\mathbb{B}_2, p_j\big)\big)}{\sum_{p_j \in \mathbb{P}} \exp\big(\epsilon Q\big(\mathbb{B}_1, p_j\big)\big)} \\
&\leq \frac{\exp(\epsilon Q(\mathbb{B}_2, p) + \epsilon \Delta)}{\exp(\epsilon Q(\mathbb{B}_2, p))} \times \frac{\sum_{p_j \in \mathbb{P}} \exp\big(\epsilon Q\big(\mathbb{B}_1, p_j\big) + \epsilon \Delta\big)}{\sum_{p_j \in \mathbb{P}} \exp\big(\epsilon Q\big(\mathbb{B}_1, p_j\big)\big)} \\
&\leq \exp(2\epsilon \Delta).
\end{aligned} \tag{9}$$

Here, Δ means the Lipschitz constant introduced in Definition 3. In our designed algorithm, the largest difference between the revenue values $Q(\mathbb{B}_1, p)$ and $Q(\mathbb{B}_2, p)$ for $\forall p \in \mathbb{P}$ and any two adjacent bid sets $\mathbb{B}_1$ and $\mathbb{B}_2$, i.e., Δ, is $\Delta = p \times (|\widetilde{\mathbb{B}_1}(p)| - |\widetilde{\mathbb{B}_2}(p)|)$. Since $\mathbb{B}_1$ and $\mathbb{B}_2$ differ in one bid, and the price values in $\mathbb{P}$ are mapped into (0, 1], we have

$$\Delta = p \times \big(\big|\widetilde{\mathbb{B}_1}(p)\big| - \big|\widetilde{\mathbb{B}_2}(p)\big|\big) \leq p \leq 1. \tag{10}$$

Based on (9) and (10), we get

$$\Pr\big[M(\mathbb{B}_1) = p\big] \leq \exp(2\epsilon) \times \Pr\big[M(\mathbb{B}_2) = p\big]. \tag{11}$$

Thus, our algorithm guarantees 2ϵ-differential privacy. ■

Here, 2ϵ-differential privacy can lead to a relaxation of truthfulness, in other words, the incentive to lie for each bidder is nonzero but tightly constrained. Next, we prove the approximate truthfulness as follows.

Theorem 2: The proposed algorithm is $(e^2 - 1)\epsilon$-truthful.

Proof: For simplicity of the following descriptions, we first let $E[u_i(b_i, \mathbb{B}_{-i}, v_i, \mathbb{P})]$ denote the expected utility of bidder i when it bids b_i. If $b_i \neq_i$, we have two following cases.

1) When $b_i \leq_i$, we have

$$\begin{aligned}
&E\big[u_i(b_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big] \\
&= \sum\nolimits_{p \in \mathbb{P}} \Pr\big[M(b_i, \mathbb{B}_{-i}) = p\big] \times u_i(b_i, \mathbb{B}_{-i}, v_i, p) \\
&\leq \sum\nolimits_{p \in \mathbb{P}} \exp(2\epsilon) \Pr\big[M(v_i, \mathbb{B}_{-i}) = p\big] \times u_i(v_i, \mathbb{B}_{-i}, v_i, p) \\
&= \exp(2\epsilon) \times E\big[u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big]
\end{aligned} \tag{12}$$

where $\Pr[M(b_i, \mathbb{B}_{-i}) = p]$ means the probability of selecting the price p based on the bid b_i and the set of others' bids $\mathbb{B}_{-i}$, and $u_i(b_i, \mathbb{B}_{-i}, v_i, p)$ denotes the utility defined in (1).

Since ϵ is a small constant (less than 1), we have $\exp(2\epsilon) \leq 1 + (e^2 - 1)\epsilon$, where e (≈ 2.72) is the base of the natural logarithm. Moreover, due to $p, v_i, b_i \in (0, 1]$, we have $0 \leq [u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})] \leq 1$. Then, we continue (12)

$$\begin{aligned}
&\exp(2\epsilon) \times E\big[u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big] \\
&\leq \Big(1 + \Big(e^2 - 1\Big)\epsilon\Big) \times E\big[u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big] \\
&\leq E\big[u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big] + \Big(e^2 - 1\Big)\epsilon.
\end{aligned} \tag{13}$$

By combining (12) and (13), the theorem holds.

2) When $b_i > v_i$, if the bidder i loses, its utility is 0. However, if the bidder wins, i.e., $b_i \geq p$, we have two following subcases: 1) if $v_i < p \leq_i$, we get the utility (i.e., the value of $v_i - p$) is negative and 2) if $p \leq v_i \leq b_i$,

the utility equals to $v_i - p$. More specifically,

$$\begin{aligned} &E\big[u_i(b_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big] \\ &\quad \le \sum\nolimits_{p\in\mathbb{P}\wedge(p\le v_i)} \Pr\big[M(b_i, \mathbb{B}_{-i}) = p\big] \times u_i(b_i, \mathbb{B}_{-i}, v_i, p) \\ &\quad = \sum\nolimits_{p\in\mathbb{P}\wedge(p\le v_i)} \Pr\big[M(v_i, \mathbb{B}_{-i}) = p\big] \times u_i(v_i, \mathbb{B}_{-i}, v_i, p) \\ &\quad \le E\big[u_i(v_i, \mathbb{B}_{-i}, v_i, \mathbb{P})\big]. \end{aligned} \tag{14}$$

According to the above cases, we have that the false bid reported by a bidder will lead his utility to nonpositive. This indicates that the data pricing algorithm can achieve $(e^2 - 1)\epsilon$-truthfulness. ∎

Next, by letting opt denote $\max_{p\in\mathbb{P}} Q(\mathbb{B}, p) = \max_{p\in\mathbb{P}} p \cdot |\widetilde{\mathbb{B}}(p)|$, we give the approximation ratio of achieved revenues in the following theorem.

Theorem 3: The proposed data pricing algorithm has an expected revenue of at least $\text{opt} - 3\ln(e + \epsilon\text{opt}|\mathbb{P}|)/\epsilon$.

Proof: Assume $S_t = \{p \mid Q(\mathbb{B}, p) > \text{opt} - t\}$ and $\overline{S}_{2t} = \{p \mid Q(\mathbb{B}, p) \le \text{opt} - 2t\}$. Then, after taking the definition of the exponential mechanism into consideration, the relationship between the probability $\Pr[p \in \overline{S}_{2t}]$ and $\Pr[p \in S_t]$ satisfies

$$\begin{aligned} &\Pr\big[p \in \overline{S}_{2t}\big]/\Pr\big[p \in S_t\big] \\ &\quad = \frac{\sum_{p\in\overline{S}_{2t}} \exp(\epsilon Q(\mathbb{B}, p))/\sum_{p\in\mathbb{P}} \exp(\epsilon Q(\mathbb{B}, p))}{\sum_{p\in S_t} \exp(\epsilon Q(\mathbb{B}, p))/\sum_{p\in\mathbb{P}} \exp(\epsilon Q(\mathbb{B}, p))} \\ &\quad = \frac{\sum_{p\in\overline{S}_{2t}} \exp(\epsilon Q(\mathbb{B}, p))}{\sum_{p\in S_t} \exp(\epsilon Q(\mathbb{B}, p))} < \frac{\exp(\epsilon(\text{opt} - 2t))|\overline{S}_{2t}|}{\exp(\epsilon(\text{opt} - t))|S_t|} \\ &\quad \le \exp(-\epsilon t)|\mathbb{P}|/|S_t|. \end{aligned} \tag{15}$$

Then, we can easily prove that $\Pr[p \in \overline{S}_{2t}] \le (|\mathbb{P}|\exp(-\epsilon t))/|S_t|$ based on the above results. This is because $\Pr[p \in S_t] \le 1$. This means that, we select price $p \in \mathbb{P}$ which can generate at least $\text{opt} - 2t$ revenue with the probability of at least $1 - (|\mathbb{P}|\exp(-\epsilon t))/|S_t|$. Next, we focus on how to select the suitable value of t ($\ge 1/\epsilon$), so that the probability is at least $1 - t/\text{opt}$. When we let t satisfy $t \ge \ln(|\mathbb{P}|\text{opt}/(t|S_t|))/\epsilon$, we have the following inequalities:

$$\begin{aligned} &1 - (|\mathbb{P}|\exp(-\epsilon t))/|S_t| \\ &\quad \le 1 - (|\mathbb{P}|\exp(-\epsilon(\ln(|\mathbb{P}|\text{opt}/(t|S_t|))/\epsilon)))/|S_t| \\ &\quad = 1 - t/\text{opt}. \end{aligned} \tag{16}$$

This indicates that our algorithm can generate at least $\text{opt} - 2t$ revenue with the probability $1 - t/\text{opt}$. Next, the expected revenues, denoted as $\Pr[\text{rev}(M)]$, satisfy

$$\Pr[\text{rev}(M)] \ge (\text{opt} - 2t) \times (1 - t/\text{opt}) > \text{opt} - 3t. \tag{17}$$

For $t \ge \ln(|\mathbb{P}|\text{opt}/(t|S_t|))/\epsilon$ and $t \ge 1/\epsilon$, we have

$$\begin{aligned} &\ln(|\mathbb{P}|\text{opt}/(t|S_t|))/\epsilon < \ln(|\mathbb{P}|\text{opt}/(t))/\epsilon \\ &\quad < \ln(e + |\mathbb{P}|\text{opt}/t)/\epsilon < \ln(e + \epsilon|\mathbb{P}|\text{opt})/\epsilon. \end{aligned} \tag{18}$$

By letting $t = \ln(e + \epsilon|\mathbb{P}|\text{opt})/\epsilon$ where $t \ge 1/\epsilon$, we get

$$\Pr[\text{rev}(M)] \ge -3t \ge -3\ln(e + \epsilon|\mathbb{P}|\text{opt})/\epsilon. \tag{19}$$

The theorem holds. ∎

In fact, the work [2] based on machine learning for an absolute-truthfulness incentive mechanism is proved to achieve $\text{opt} - O(\sqrt{\text{opt}})$ revenue in expectation. This means that the proposed differentially private auction-based data pricing algorithm can balance the maximum revenues and the approximate truthfulness efficiently.

IV. DPDT: Data Collection Algorithm

In this section, we design a differentially private data collection algorithm. We first introduce the basic idea and then present the detailed algorithm. Finally, we analyze the privacy and performance of the proposed algorithm.

A. Basic Idea

To ensure that the data to be collected is kept private against crowd workers, we first adopt the perturbation method to hide the true data collection tasks and then use the exponential mechanism to select a minimum number of workers (according to the calculated probability distribution). Concretely, the platform first generates some noisy tasks and then adds them into true tasks. Next, the platform publishes all tasks, including true and noisy tasks, to crowd workers. Each worker selects some tasks which she/he is willing to perform, and then sends the results to the platform. At last, the platform aims to select a minimum number of workers to cover the private true task set. Although the crowd workers receive all (true and noisy) data collection tasks, they cannot specifically indicate these true tasks. At the same time, by adopting the exponential mechanism in the worker recruitment process, the true data collection tasks can be protected efficiently.

More specifically, given the total data collection tasks $\mathbb{U}$ consisting of true tasks $\mathbb{R}$ and noisy tasks $\mathbb{T}$, i.e., $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$, the set $\mathbb{G}_k$ of tasks which each worker $k \in \mathbb{W}$ is willing to perform, and the collection of $\mathbb{G}_k$, i.e., $\Omega = \{\mathbb{G}_k | k \in \mathbb{W} \wedge \mathbb{G}_k \subseteq \mathbb{U}\}$, we select a minimum number of workers so that we can cover a private subset $\mathbb{R} \subseteq \mathbb{U}$. The whole worker recruitment process contains multiple rounds of iterations. Since only one element (without loss of generality, denoted by $\mathbb{G}_k$) in Ω can be selected in each round, the uncovered tasks in $\mathbb{R}$ are updated by $\mathbb{R} - \mathbb{G}_k$. Moreover, the collection Ω is also updated by $\Omega - \{\mathbb{G}_k\}$. The probability of selecting a subset of $\mathbb{U}$ (e.g., $\mathbb{G}_k$) in Ω is proportional to the number of the intersecting elements between $\mathbb{G}_k$ and $\mathbb{R}$, that is,

$$\Pr[\mathbb{G}_k \in \Omega] = \frac{\exp\big(\epsilon'|\mathbb{G}_k \cap \mathbb{R}|\big)}{\sum_{\mathbb{G}\in\Omega} \exp(\epsilon'|\mathbb{G} \cap \mathbb{R}|)} \tag{20}$$

where $\epsilon' = \epsilon/(2\ln(e/\delta))$ is a small constant.

B. Detailed Algorithm

The differentially private data collection algorithm is shown in Algorithm 2, which mainly consists of five phases. In the first three phases, the platform first generates a noisy task set $\mathbb{T}$ and then adds $\mathbb{T}$ into $\mathbb{R}$. Here, the geographical distributions of $\mathbb{T}$ and $\mathbb{R}$ are similar and the cardinalities of $\mathbb{T}$ and $\mathbb{R}$ are on the same order of magnitude. Then, the platform publishes all tasks $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$ to crowd workers. Next, each worker responds to the platform with the tasks that she/he is willing to perform. In the fourth phase, the platform focuses on how

Algorithm 2 Differentially Private Data Collection Algorithm

Require: $\mathbb{R}$, $\mathbb{W}$, ϵ, δ

Ensure: Φ and Ψ

Phase 1: platform generates noisy data collection tasks $\mathbb{T}$;

// The distributions of $\mathbb{T}$ and $\mathbb{R}$ are similar and the cardinalities of $\mathbb{T}$ and $\mathbb{R}$ are on the same order of magnitude;

Phase 2: platform publishes $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$ to workers $\mathbb{W}$;

Phase 3: each worker $k \in \mathbb{W}$ submits $\mathbb{G}_k$ to the platform, where $\Omega = \{\mathbb{G}_k | k \in \mathbb{W} \wedge \mathbb{G}_k \subseteq \mathbb{U}\}$;

Phase 4: platform privately selects suitable workers;

1: Initialize $\Phi \leftarrow \phi$, $i \leftarrow 1$, $\mathbb{R}_i \leftarrow \mathbb{R}$, $\Omega_i \leftarrow \Omega$, $\epsilon' \leftarrow \frac{\epsilon}{2\ln(e/\delta)}$;
2: **for** $i = 1, 2, \cdots, |\mathbb{W}|$ **do**
3: pick a set $\mathbb{G}_k$ from Ω_i with the following probability $Pr[\mathbb{G}_k \in \Omega_i] = \frac{\exp(\epsilon'|\mathbb{G}_k \cap \mathbb{R}_i|)}{\sum_{\mathbb{G} \in \Omega_i} \exp(\epsilon'|\mathbb{G} \cap \mathbb{R}_i|)}$;
4: $\Phi = \Phi \cup \{k\}$ and $\Psi = \Psi \cup \{\mathbb{G}_k\}$;
5: $\mathbb{R}_{i+1} \leftarrow \mathbb{R}_i - \mathbb{G}_k$, $\Omega_{i+1} \leftarrow \Omega_i - \{\mathbb{G}_k\}$;
6: **if** $\mathbb{R}_{i+1} = \phi$ **then**
7: break;
8: **end if**
9: **end for**
10: **output** the selected worker set Φ and Ψ;

Phase 5: the workers in Φ collect data and send it to platform;

to select a minimum number of workers to collect data while protecting data privacy. More specifically, the platform first initializes several parameters, such as the selected worker set Φ, the new differential privacy budget ϵ', etc., in step 1. Then, the platform selects a set $\mathbb{G}_k$ from Ω_i with the probability $\exp(\epsilon'|\mathbb{G}_k \cap \mathbb{R}_i|)/\sum_{\mathbb{G} \in \Omega_i} \exp(\epsilon'|\mathbb{G} \cap \mathbb{R}_i|)$, and further adds k and $\mathbb{G}_k$ into Φ and Ψ, respectively, in steps 2–4. In step 5, the remaining elements in $\mathbb{R}$ are updated as $\mathbb{R}_{i+1} \leftarrow \mathbb{R}_i - \mathbb{G}_k$, and Ω is updated as $\Omega_{i+1} \leftarrow \Omega_i - \{\mathbb{G}_k\}$. When no task exists in $\mathbb{R}$, i.e., $\mathbb{R} = \phi$, the algorithm terminates and outputs the selected worker set Φ as well as Ψ, in steps 6–8. In the fifth phase, the selected workers are required to collect the corresponding crowd-sensed data and then send the results to the platform.

By analyzing Algorithm 2, we show that the algorithmic procedures are polynomial-time. Specifically, the computational overhead of Algorithm 2 is dominated by step 3, denoted by $O(|\mathbb{W}|^2 \cdot |\mathbb{R}| \cdot |\mathbb{G}|)$, where $|\mathbb{G}| = \max_{k \in \mathbb{W}} |\mathbb{G}_k|$.

C. Performance Analysis

Now, we analyze the differential privacy level and the approximation ratio. We first let $|\mathbb{U}| = x$ and $|\Omega| = |\mathbb{W}| = y$ for simplicity. Then, we get the following theorem.

Theorem 4: For any $\delta < 1/e$ and $\epsilon \in (0, 1)$, the proposed algorithm can preserve δ-approximate ϵ-differential privacy.

Proof: For simplicity of the following descriptions, we use $\mathbb{R}$ and $\mathbb{R}^+$ to denote two private task set instances, in which $\mathbb{R}$ and $\mathbb{R}^+$ differ in one element o. Also, we let Ω^o denote the collection of task sets containing o. Consider that an output permutation is denoted by π. After the first $i-1$ sets in π have been added into the cover, we use $s_{i,k}(\mathbb{R})$ to denote the number of valid elements in the set $\mathbb{G}_k$. "Valid elements" here mean those in the set $\mathbb{R}_{i-1} \cap \mathbb{G}_k$. Based on this, we have

$$\begin{aligned}
&\Pr[M(\mathbb{R}) = \pi]/\Pr\big[M\big(\mathbb{R}^+\big) = \pi\big] \\
&= \prod_{i=1}^{x} \left(\frac{\exp\big(\epsilon' s_{i,\pi_i}(\mathbb{R})\big)/\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}(\mathbb{R})\big)}{\exp\big(\epsilon' s_{i,\pi_i}\big(\mathbb{R}^+\big)\big)/\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}\big(\mathbb{R}^+\big)\big)} \right) \\
&= \frac{\exp\big(\epsilon' s_{t,\pi_t}(\mathbb{R})\big)}{\exp\big(\epsilon' s_{t,\pi_t}\big(\mathbb{R}^+\big)\big)} \times \prod_{i=1}^{t} \left(\frac{\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}\big(\mathbb{R}^+\big)\big)}{\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}(\mathbb{R})\big)} \right) \quad (21)
\end{aligned}$$

where t indicates that $\mathbb{G}_{\pi_t}$ is the first set containing the element o to be added into the permutation π. That is to say, after the t-th iteration, the remaining elements in $\mathbb{R}$ and $\mathbb{R}^+$ are identical. In (21), except for t-th term, all other terms in the numerators and denominators cancel each other out. This is because that all the corresponding set sizes are equal. For the relationship of $\mathbb{R}$ and $\mathbb{R}^+$, we have two cases.

Case 1: $\mathbb{R} - \mathbb{R}^+ = \{o\}$. We have that the first term of (21) is $\exp(\epsilon')$ and the other term in the product is at most 1. So, (21) is less than $\exp(\epsilon')$.

Case 2: $\mathbb{R}^+ - \mathbb{R} = \{o\}$. Here, we get that the first term of (21) is $\exp(-\epsilon') < 1$. In this case, each set in Ω^o for instance $\mathbb{R}^+$ is larger by 1 than that for $\mathbb{R}$, while other sets are identical. Based on this, we have

$$\begin{aligned}
&\Pr[M(\mathbb{R}) = \pi]/\Pr\big[M\big(\mathbb{R}^+\big) = \pi\big] \\
&\le \prod_{i=1}^{t} \frac{\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}(\mathbb{R})\big) + \big(\exp(\epsilon') - 1\big) \sum_{\mathbb{G}_k \in \Omega^o} \exp\big(\epsilon' s_{i,k}(\mathbb{R})\big)}{\sum_{k \in \mathbb{W}} \exp\big(\epsilon' s_{i,k}(\mathbb{R})\big)} \\
&= \prod_{i=1}^{t} \big(1 + (\exp(\epsilon') - 1)\mathrm{pr}_i(\mathbb{R})\big) \quad (22)
\end{aligned}$$

where $\mathrm{pr}_i(\mathbb{R})$ represents the probability that a set containing the element o is chosen at ith step based on the private instance $\mathbb{R}$. Now, the previous steps have selected the sets $\mathbb{G}_{\pi_1}, \mathbb{G}_{\pi_2}, \ldots, \mathbb{G}_{\pi_{i-1}}$.

For a private task set instance $\mathbb{R}$, we say that an output is α − good if $\sum_i \mathrm{pr}_i(\mathbb{R}) \le \alpha$. Otherwise, we call the output α − bad when $\sum_i \mathrm{pr}_i(\mathbb{R}) > \alpha$. Thus, we first consider the case where the permutation π is $(\ln \delta^{-1})$ − good. Considering the definition of t, we have

$$\sum_{i=1}^{t-1} \mathrm{pr}_i(\mathbb{R}) \le \ln \delta^{-1}. \quad (23)$$

Based on this, we continue (22) and have

$$\begin{aligned}
&\Pr[M(\mathbb{R}) = \pi]/\Pr\big[M\big(\mathbb{R}^+\big) = \pi\big] \\
&\le \prod_{i=1}^{t} \big(\exp\big(\exp(\epsilon') - 1\big)\mathrm{pr}_i(\mathbb{R})\big) \le \exp\left(2\epsilon' \sum_{i=1}^{t} \mathrm{pr}_i(\mathbb{R})\right) \\
&\le \exp\Big(2\epsilon'\Big(\ln \delta^{-1} + \mathrm{pr}_t(\mathbb{R})\Big)\Big) \le \exp\Big(2\epsilon'(\ln \delta^{-1} + 1)\Big) \\
&= \exp\Big(2(\epsilon/(2\ln(e/\delta)))(\ln \delta^{-1} + 1)\Big) = \exp(\epsilon). \quad (24)
\end{aligned}$$

Next, for any set Ψ of outcomes, we get

$$\begin{aligned}
\Pr[M(\mathbb{R}) \in \Psi] &= \sum_{\pi \in \Psi} \Pr[M(\mathbb{R}) = \pi] \\
&= \sum_{\pi \in \Psi \wedge \pi \in \text{good}} \Pr[M(\mathbb{R}) = \pi]
\end{aligned}$$

$$+ \sum_{\pi \in \Psi \wedge \pi \in \text{bad}} \Pr[M(\mathbb{R}) = \pi]$$
$$\leq \sum_{\pi \in \Psi \wedge \pi \in \text{good}} \exp(\epsilon) \Pr\left[M\left(\mathbb{R}^{+}\right) = \pi\right] + \delta$$
$$\leq \exp(\epsilon) \Pr\left[M\left(\mathbb{R}^{+}\right) = \pi\right] + \delta \quad (25)$$

where $\pi \in$ good and $\pi \in$ bad denote the sets in which π is $(\ln \delta^{-1})-$ good and $(\ln \delta^{-1})-$ bad, respectively.

As a result, the proposed algorithm preserves δ-approximate ϵ-differential privacy for $\epsilon \in (0, 1)$ and $\delta < 1/e$. The theorem holds. ■

Intuitively, when we increase the size of the noisy task set $\mathbb{T}$, we will improve the differential privacy level. In fact, the intuition is accurate here. In the above proof, we use $\text{pr}_i(\mathbb{R})$ to denote the probability that a set containing the element o is chosen at ith step based on $\mathbb{R}$. Here, $\text{pr}_i(\mathbb{R})$ will decrease with the increase of the noisy task set $\mathbb{T}$. This is because that the true tasks in $\mathbb{G}_k$ ($k \in \mathbb{W}$) decrease when $|\mathbb{G}_k|$ remains almost unchanged. Then, we have $\sum_{i=1}^{t-1} \text{pr}_i(\mathbb{R}) \leq -\ln \delta^*$ where δ^* is a small constant and $\delta^* \leq \delta$. Based on this, we have $\Pr[M(\mathbb{R}) \in \Psi] = \sum_{\pi \in \Psi} \Pr[M(\mathbb{R}) = \pi] \leq \exp(\epsilon)\Pr[M(\mathbb{R}^{+}) = \pi] + \delta^*$ here. Thus, we conclude that we can achieve the higher differential privacy level when given the larger noisy task set $\mathbb{T}$.

Next, we analyze the approximation ratio of the proposed algorithm. Before the ith iteration, according to $|\mathbb{U}| = x$ and $|\Omega| = |\mathbb{W}| = y$, we let $y_i = y - i + 1$ and $x_i = |\mathbb{R}_i|$ denote the numbers of the remaining sets and the remaining elements. Moreover, let $L_i = \max_{\mathbb{G} \in \Omega} |\mathbb{G} \cap \mathbb{R}_i|$ denote the largest number of intersecting elements covered by any set in Ω. By a standard argument [9], any algorithm that always picks the set of size $L_i/2$ is an $O(\ln n)$-approximation algorithm. Thus, we have the following theorem.

Theorem 5: The proposed differentially private data collection algorithm can achieve an expected approximation ratio of $O(\ln |\mathbb{U}| + [\ln(|\mathbb{W}| \ln(e/\delta))]/\epsilon)$.

Proof: At least one set in Ω contains L_i elements. According to the exponential mechanism used in differential privacy [9], [11], we get that the probability of selecting a set covering fewer than $L_i - 3\ln(y/\epsilon)$ elements is at most $1/y^2$. In particular, for $L_i > 6\ln(y/\epsilon)$, we always select sets that cover at least $L_i/2$ elements with probability at least $1 - 1/y$. Therefore, we use no more than $O(\text{OPT} \ln |\mathbb{U}|)$ sets where OPT means the optimal solution. While $L_i \leq 6\ln(y/\epsilon)$, the number of remaining elements $\mathbb{R}_i$ is at most $\text{OPT} \times_i$. Thus, any permutation consumes at most an additional $O(\text{OPT} \ln y/\epsilon')$. By combining these two cases, we get that the maximum number of selected task sets is denoted as $O(\text{OPT}(\ln x + \ln y/\epsilon'))$. That is, we have $(|\Phi|/\text{OPT}) \leq O(\ln x + \ln[y/\epsilon'])$. Based on this, we conclude that the expected approximation ratio is $O(\ln x + (\ln y)/\epsilon')$, i.e., $O(\ln |\mathbb{U}| + \ln(|\mathbb{W}| \ln(e/\delta))/\epsilon)$. ■

According to Theorem 5, we get that the expected approximation ratio of the data collection algorithm is denoted as $O(\ln |\mathbb{R} \cup \mathbb{T}| + \ln(|\mathbb{W}| \ln(e/\delta))/\epsilon)$. When we add more noisy tasks into the task publishing process (i.e., increasing $|\mathbb{T}|$), the expected approximation ratio increases accordingly. In order to balance the achieved differential privacy level and the expected approximation ratio, we generally set the size of the noisy task set $\mathbb{T}$ to that of the true task set $\mathbb{R}$.

TABLE II
SIMULATION SETTINGS

Parameter name	default	range
number of consumers, $\lvert\mathbb{N}\rvert$	200	$100 - 1000$
differential privacy budget, ϵ	0.1	$0.1 - 0.5$
number of data collection tasks, $\lvert\mathbb{R}\rvert$	200	$100 - 600$
parameter, λ	1	$0.5 - 3$
parameter, η	0.2	$0.1 - 0.5$

V. PERFORMANCE EVALUATION

In this section, we evaluate the performance of DPDT with extensive simulations. We evaluate the data pricing algorithm and the data collection algorithm in two parts.

A. Evaluation on Data Pricing Algorithm

We first introduce the methodology, and then present the evaluation results, including privacy and revenues.

1) Methodology: Since the data pricing algorithm only involves data consumers and platform, we omitted crowd workers here. More specifically, we varied the number of data consumers (i.e., bidders) from 100 to 1000. For each bidder, we generated its true valuation, which was uniformly distributed over (0, 1]. Here, we first assumed that the true valuations $\mathbb{V}$ are the same as the submitted bids $\mathbb{B}$. Then, we let one individual's bid be different from its true valuation to evaluate the expected utility of the bidder. We also created the price set $\mathbb{P}$ which contains all generated valuations and bids in (0, 1]. The cardinality of the price set $\mathbb{P}$ is the number of different true valuations and bids values in (0, 1], i.e., $|\mathbb{V} \cup \mathbb{B}|$. In addition, we set the differential privacy constant ϵ from 0.1 to 0.5. The default and range for some parameters are displayed in Table II.

Since our data pricing model involves the auction (approximate truthfulness) and privacy (non-numeric outputs) simultaneously, there are no existing data trading algorithms that can be applied to our model. We implemented the *optimal* algorithm without privacy preservation for comparison. In the optimal algorithm, the price in $\mathbb{P}$ which can maximize the revenue of the platform is selected as the single payment. In addition, in order to evaluate the impact of differential privacy constant ϵ on the performance of the proposed algorithm, we ran the algorithm with different ϵ values. All the evaluation results under the same settings are averaged over 1000 times.

To evaluate the performance of the differentially private auction-based data pricing algorithm, we used four evaluation metrics: 1) revenues; 2) expected utility of one bidder; 3) probability distribution; and 4) privacy leakage. The revenues mean the total payment received from winning bidders. The expected utility of one bidder indicates the achieved utility when the bidder changes its bid, while other bidders remain unchanged. As introduced earlier, when the outcome of the probability distribution over prices only changes slightly if anyone bidder changes its submitted bid in a mechanism, we say that this mechanism can guarantee good privacy. Thus, we also used the probability distribution as an evaluation metric. Finally, we define privacy leakage.

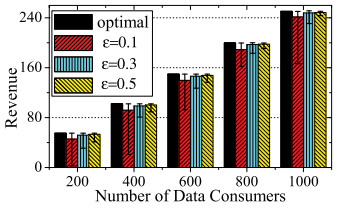


Fig. 2. Revenues versus $|\mathbb{N}|$.

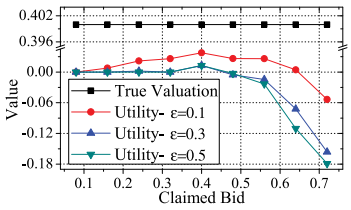


Fig. 3. Approximate truthfulness.

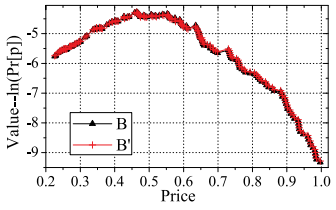


Fig. 4. Probability distribution.

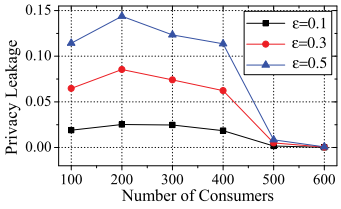


Fig. 5. Privacy leakage.

Definition 5 (Privacy Leakage): For two input profiles D and D' which differ in only one bid in a mechanism, we use S and S' to denote the corresponding probability distributions over a price set $\mathbb{P}$. We let the average of the absolute differences between the logarithmic probabilities of the two distributions denote the privacy leakage [42], that is,

$$1/|\mathbb{P}| \sum_{p_j \in \mathbb{P}} \left| \ln \Pr\left[S_j\right] - \ln \Pr\left[S'_j\right] \right|. \tag{26}$$

2) Evaluation Results: We display the evaluation results.

Evaluation of Revenues: We first evaluate the effects of the number of consumers $|\mathbb{N}|$ and the differential privacy constant ϵ on the expected revenues. The results are shown in Fig. 2. We find that our algorithm achieves high revenues which are very close to the optimal results. Moreover, the smaller the constant ϵ is, the smaller the expected revenue is. Also, a smaller ϵ value results in a wider distribution of the expected revenue. The revenues rise continuously along with the increase of the number of consumers.

Evaluation of Utility: We verify the approximate truthfulness of the data pricing algorithm by randomly picking a bidder and allowing it to submit a bid that is different from its true valuation. We observe that a bid higher than the true valuation may make the expected utility of this bidder negative, while a bid lower than the true valuation will yield the expected utility close to 0, as shown in Fig. 3. Especially, the larger ϵ is, the smaller the expected utility is.

Evaluation of Probability Distribution: We evaluate the probability distribution for two bid profiles which differ in only one bid. For a finer observation, we use the logarithmic function $\ln(\cdot)$ to amplify the probability values. The results are presented in Fig. 4. We discover that the probability distributions for the two profiles over the price set are almost identical. This means a good differential privacy.

Evaluation of Privacy Leakage: Finally, we verify the privacy leakage of the differentially private data pricing algorithm. The results show that the maximum privacy leakage value is less than 0.15 where $\epsilon = 0.5$, as displayed in Fig. 5. We observe that the privacy leakage values rise along with the increase of ϵ. These results are consistent with our theoretical analysis.

B. Evaluation on Data Collection Algorithm

1) Methodology: First, we vary the cardinality of the set $\mathbb{R}$ from 100 to 600 with the step as 100. Then, we set the number of crowd workers (i.e., $|\mathbb{W}|$) as 200. When generating noisy data collection tasks $\mathbb{T}$, we set the cardinality of $\mathbb{T}$ to be $\lambda \cdot |\mathbb{R}|$ where λ is selected from $\{0.5, 1, 1.5, 2, 2.5, 3\}$. For each crowd worker $k \in \mathbb{W}$, the cardinality of the subset of $\mathbb{U} = \mathbb{R} \cup \mathbb{T}$ that it claims to perform (i.e., $\mathbb{G}_k$) is set as $|\mathbb{G}_k| \leq \eta |\mathbb{U}|$, in which η is selected in $\{0.1, 0.2, 0.3, 0.4, 0.5\}$. Note that we will slightly control the generation of $\mathbb{G}_k$, to ensure that at least the total workers can cover the private set $\mathbb{R}$. This is reasonable because we can add more crowd workers until $\mathbb{R}$ can be covered.

We design and implement a *greedy* algorithm without privacy preservation for comparison. The greedy algorithm always selects the set $\mathbb{G}_k$, which covers the maximum number of the remaining tasks in $\mathbb{R}$, i.e., $\max_{\mathbb{G}_k \in \Omega} |\mathbb{G}_k \cap \mathbb{R}|$. Ω and $\mathbb{R}$ are updated in time after $\mathbb{G}_k$ is selected. Moreover, we execute our differentially private data collection algorithm 200 times under different privacy levels (i.e., $\epsilon \in \{0.1, 0.3, 0.5\}$). To evaluate the performance of the proposed algorithm, we use the following metrics: the cardinality of the solution set, i.e., $|\Phi|$, and the probability distribution. The small $|\Phi|$ value means the small recruitment cost. The probability distribution for the private set $\mathbb{R}$ indicates the possibility of each worker being selected in the beginning.

2) Evaluation Results: We exhibit the results as follows.

Evaluation of $|\Phi|$: We first evaluate the performance of $|\Phi|$ when we change the number of true data collection tasks (i.e., $|\mathbb{R}|$), as shown in Fig. 6. We find that the greedy algorithm without privacy-preservation outputs the best and error-free results under the same settings. This is because no perturbation exists in the greedy algorithm. In our algorithm, the smaller the privacy level ϵ is, the worse the result of $|\Phi|$ is. This is because we cannot achieve good performance and satisfactory privacy simultaneously. Next, we evaluate the effects of the

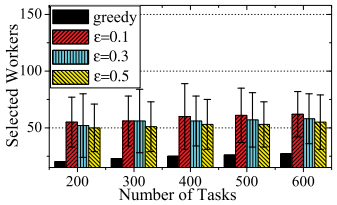


Fig. 6. $|\Phi|$ versus $|\mathbb{R}|$.

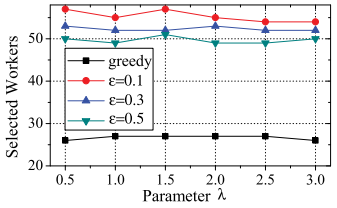


Fig. 7. $|\Phi|$ versus parameter λ.

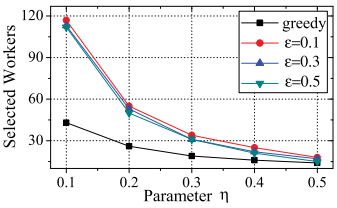


Fig. 8. $|\Phi|$ versus parameter η.

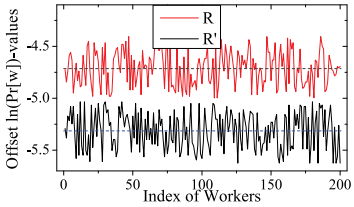


Fig. 9. Probability distribution.

parameters λ and η. The number of selected workers (i.e., $|\Phi|$) remains almost unchanged along with the increase of λ, as shown in Fig. 7, while $|\Phi|$ decreases by increasing the value of η, as shown in Fig. 8. In both Figs. 7 and 8, our algorithm with larger privacy level ϵ obtains better performance. These observations are consistent with our theoretical analysis.

Evaluation of Probability Distribution: At the beginning of this algorithm, we compare the probability distributions for two private sets $\mathbb{R}$ and $\mathbb{R}'$ which differ in one element, as shown in Fig. 9. The probabilities here are amplified by the logarithmic function $\ln(\cdot)$. In order to distinguish the amplified probability distributions, we make the probabilities based on $\mathbb{R}'$ plus a constant 0.6. We observe that the two probability distributions have a similar trend, and the average logarithmic function values for $\mathbb{R}$ and $\mathbb{R}'$ are -5.3141 and -5.3119, respectively, which are almost identical. These simulations validate our theoretical analysis results.

VI. Related Work

In this article, we focus on the problem of designing a privacy-preserving crowd-sensed data trading mechanism. So far, there has been much research on the crowd-sensed data trading problem, such as [3], [7], [10], [12], [15], [23], [36], [37], [39], and [41], and the mobile crowdsensing problem, such as [1], [13], [14], [17]–[20], [27]–[30], [32], [38], and [40].

More specifically, Zheng *et al.* [39] proposed a data acquisition scheme for crowd-sensed data markets. Yu *et al.* [36] introduced a prospect theory model from behavioral economics to understand the users' realistic trading behaviors, and then design an algorithm to help estimate the users' risk preference and dynamically provide trading recommendations. Cao *et al.* [3] proposed an iterative auction mechanism for the data trading problem, which can guide multiple selfish data agents (including data owners, collectors, and users) to trade data efficiently in terms of social welfare. Jung *et al.* [15] studied the responsibilities of the consumers in the dataset trading, and then design the accountable protocols such that the book-keeping ability and accountability against dishonest consumers are achieved throughout the dataset transactions. After considering the multiple task initiators and participants in the mobile crowdsensing, He *et al.* [10] proposed the concept of "Walrasian Equilibrium," based on which they find the Pareto optimal task allocation for initiators.

However, the above works rarely consider the privacy-preserving issues in the data trading market. In particular, only a few works have involved the identity privacy of consumers or data privacy in the data trading process. Gao *et al.* [7] integrated homomorphic encryption technique into the auction-based big data trading to protect the bid privacy. Zhang *et al.* [37] designed a privacy-preserving crowdsourcing-based image dataset purchasing framework, in which buyers can purchase the image datasets that meet their quality requirements. Niu *et al.* [23] adopted homomorphic encryption and identity-based signature to design a truthful and privacy-preserving data trading mechanism. Nevertheless, the homomorphic encryption will result in a huge computation or communication overhead that is unacceptable to data consumers. Li *et al.* [16] studied the location-sharing privacy leakage problem in mobile social networks, while To *et al.* [28] designed a differentially private geocast-based framework to protect the location privacy of workers in mobile crowdsensing. Also, Li and Cao [17], Lin *et al.* [19], and Wang *et al.* [30] proposed some privacy-aware incentive mechanisms for mobile crowdsensing. Moreover, Jin *et al.* [13] devised an incentive mechanism for privacy-aware data aggregation, while Li *et al.* [18] designed an efficient mechanism for privacy-preserving truth discovery in mobile crowdsensing systems.

Different from the aforesaid works, we design a privacy-preserving crowd-sensed data trading mechanism, including data pricing and data collection. We consider the identity privacy of consumers and data collection task privacy in the data pricing and data collection, respectively. To this end, we design a differentially private auction-based data pricing algorithm, which achieves a good approximation to the maximum revenue, preserves the identity privacy of data consumers, and at the same time guarantees an approximate-truthfulness during the process of bidding. None of the existing privacy-preserving data trading mechanisms can achieve these goals

simultaneously. We also propose a differentially private data collection algorithm, by modeling our problem as a special set cover problem with differential privacy. Indeed, it is challenging to simultaneously achieve good performance and a satisfying level of privacy. Our proposed data collection algorithm can ensure a tight bound of the expected approximation ratio and meanwhile obtain a good level of differential privacy.

Differential privacy [11], [22], as a method to limit the disclosure of private information records in a statistical dataset, was first introduced by Dwork [4] in 2006, and it has attracted lots of attention and researches recently. In general, the Laplace mechanism and exponential mechanism are the two most commonly used differential privacy methods. The former involves adding random noise into the numeric queries so that the answers conform to the Laplace statistical distribution, while the latter is designed for the non-numeric queries and it makes high-quality outputs exponentially more likely at a rate that depends on the sensitivity of the quality score and the privacy parameter. Since our crowd-sensed data trading scenario involves a non-numeric output, we adopted the exponential mechanism to preserve the privacy in this article. Furthermore, our data pricing also deals with the tradeoff between the revenue maximization and the approximate truthfulness in the auction mechanism [22], [42]. Therefore, we combine the exponential mechanism with the approximate-incentive mechanism to devise a data pricing algorithm, instead of the simple application of differential privacy.

VII. Conclusion

In this article, we propose a differentially private crowd-sensed data trading mechanism (i.e., DPDT), which consists of a data pricing algorithm and a data collection algorithm. We take the first step to integrate the differential privacy (exponential mechanism) into the crowd-sensed data trading market to preserve identity and task privacy. The data pricing algorithm not only realizes 2ϵ-differential privacy and $(e^2-1)\epsilon$-truthfulness, but also achieves an expected revenue of at least $\text{opt} - 3\ln(e+\epsilon|\mathbb{P}|\text{opt})/\epsilon$, where opt is the optimal revenue and $\mathbb{P}$ is the set of possible prices. The data collection algorithm obtains δ-approximate ϵ-differential privacy, and meanwhile achieves an expected approximation ratio of $O(\ln|\mathbb{U}|+(\ln(|\mathbb{W}|\ln(e/\delta)))/\epsilon)$, in which $\mathbb{U}$ and $\mathbb{W}$ are the sets of total tasks and crowd workers, respectively. Extensive simulations were conducted to verify the significant performance of DPDT.

Acknowledgment

The authors would like to thank support provided by the China Scholarship Council (CSC) during a visit of G. Gao to Temple University (Grant CSC No. 201806340014).

References

[1] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, and Z. Han, "The accuracy-privacy trade-off of mobile crowdsensing," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 132–139, Jun. 2017.

[2] M.-F. Balcan, A. Blum, J. D. Hartline, and Y. Mansour, "Mechanism design via machine learning," in *Proc. 46th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, 2005, pp. 605–614.

[3] X. Cao, Y. Chen, and K. J. R. Liu, "Data trading with multiple owners, collectors, and users: An iterative auction mechanism," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 3, no. 2, pp. 268–281, Jun. 2017.

[4] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, 2008, pp. 1–19.

[5] G. Gao, M. Xiao, J. Wu, H. Huang, S. Wang, and G. Chen, "Auction-based VM allocation for deadline-sensitive tasks in distributed edge cloud," *IEEE Trans. Services Comput.*, to be published.

[6] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2982–2997, Dec. 2018.

[7] W. Gao, W. Yu, F. Liang, W. G. Hatcher, and C. Lu, "Privacy-preserving auction for big data trading using homomorphic encryption," *IEEE Trans. Netw. Sci. Eng.*, to be published.

[8] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy, and incentive provision for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 839–853, Oct. 2016.

[9] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar, "Differentially private combinatorial optimization," in *Proc. ACM-SIAM Symp. Discr. Algorithms*, 2010, pp. 1106–1125.

[10] S. He, D.-H. Shin, J. Zhang, J. Chen, and P. Lin, "An exchange market approach to mobile crowdsensing: Pricing, task allocation, and Walrasian equilibrium," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 921–934, Apr. 2017.

[11] Z. Huang and S. Kannan, "The exponential mechanism for social welfare: Private, truthful, and nearly optimal," in *Proc. Annu. Symp. Found. Comput. Sci.*, 2012, pp. 140–149.

[12] Y. Jiao, P. Wang, S. Feng, and D. Niyato, "Profit maximization mechanism and data management for data analytics services," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 2001–2014, Jun. 2018.

[13] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2019–2032, Oct. 2018.

[14] W. Jin, M. Li, L. Guoy, and L. Yang, "DPDA: A differentially private double auction scheme for mobile crowd sensing," in *Proc. Conf. Commun. Netw. Security*, 2018, pp. 1–9.

[15] T. Jung *et al.*, "AccountTrade: Accountable protocols for big data trading against dishonest consumers," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2017, pp. 1–9.

[16] H. Li, H. Zhu, S. Du, X. Liang, and X. Shen, "Privacy leakage of location sharing in mobile social networks: Attacks and defense," *IEEE Trans. Depend. Secure Comput.*, vol. 15, no. 4, pp. 646–660, Jul./Aug. 2018.

[17] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proc. IEEE 34th Int. Conf. Distrib. Comput. Syst.*, 2014, pp. 208–217.

[18] Y. Li *et al.*, "An efficient two-layer mechanism for privacy-preserving truth discovery," in *Proc. ACM SIGKDD Int. Conf. Knowl. Disc. Data Min.*, 2018, pp. 1705–1714.

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

[20] Q. Liu, G. Wang, F. Li, S. Yang, and J. Wu, "Preserving privacy with probabilistic indistinguishability in weighted social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 28, no. 5, pp. 1417–1429, May 2017.

[21] Q. Liu, G. Wang, X. Liu, T. Peng, and J. Wu, "Achieving reliable and secure services in cloud computing environments," *Comput. Elect. Eng.*, vol. 59, pp. 153–164, Apr. 2017.

[22] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. 48th Annu. IEEE Symp. Found. Comput. Sci. (FOCS)*, 2007, pp. 94–103.

[23] C. Niu, Z. Zheng, F. Wu, X. Gao, and G. Chen, "Achieving data truthfulness and privacy preservation in data markets," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 1, pp. 105–119, Jan. 2019.

[24] L. Ou, Z. Qin, S. Liao, Y. Hong, and X. Jia, "Releasing correlated trajectories: Towards high utility and optimal differential privacy," *IEEE Trans. Depend. Secure Comput.*, to be published.

[25] W. H. Qardaji and N. Li, "Recursive partitioning and summarization: A practical framework for differentially private data publishing," in *Proc. 7th ACM Symp. Inf. Comput. Commun. Security*, 2012, pp. 38–39.

[26] Z. Qin, F. Chen, Q. Wang, A. X. Liu, and Z. Qin, "Towards high performance security policy evaluation," *J. Supercomput.*, vol. 59, no. 3, pp. 1577–1595, 2012.

[27] Y. Qu *et al.*, "Posted pricing for chance constrained robust crowdsensing," *IEEE Trans. Mobile Comput.*, to be published.

[28] H. To, G. Ghinita, L. Fan, and C. Shahabi, "Differentially private location protection for worker datasets in spatial crowdsourcing," *IEEE Trans. Mobile Comput.*, vol. 16, no. 4, pp. 934–949, Apr. 2017.

[29] Q. Wang, Y. Zhang, X. Lu, Z. Wang, Z. Qin, and K. Ren, "Real-time and spatio-temporal crowd-sourced social network data publishing with differential privacy," *IEEE Trans. Depend. Secure Comput.*, vol. 15, no. 4, pp. 591–606, Jul./Aug. 2018.
[30] X. Wang, Z. Liu, X. Tian, X. Gan, Y. Guan, and X. Wang, "Incentivizing crowdsensing with location-privacy preserving," *IEEE Trans. Wireless Commun.*, vol. 16, no. 10, pp. 6940–6952, Oct. 2017.
[31] Y. Wei, Y. Zhu, H. Zhu, Q. Zhang, and G. Xue, "Truthful online double auctions for dynamic mobile crowdsourcing," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2015, pp. 2074–2082.
[32] Y. Wu, F. Li, L. Ma, Y. Xie, T. Li, and Y. Wang, "A context-aware multi-armed bandit incentive mechanism for mobile crowd sensing systems," *IEEE Internet Things J.*, to be published.
[33] M. Xiao, J. Wu, S. Zhang, and J. Yu, "Secret-sharing-based secure user recruitment protocol for mobile crowdsensing," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2017, pp. 1–9.
[34] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for *k*-anonymity location privacy," in *Proc. IEEE INFOCOM*, 2013, pp. 2994–3002.
[35] H. Yin, Z. Qin, L. Ou, and K. Li, "A query privacy-enhanced and secure search scheme over encrypted data in cloud computing," *J. Comput. Syst. Sci.*, vol. 90, pp. 14–27, Dec. 2017.
[36] J. Yu, M. H. Cheung, J. Huang, and H. V. Poor, "Mobile data trading: Behavioral economics analysis and algorithm design," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 4, pp. 994–1005, Apr. 2017.
[37] L. Zhang *et al.*, "CrowdBuy: Privacy-friendly image dataset purchasing via crowdsourcing," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2018, pp. 2735–2743.
[38] M. Zhang, L. Yang, X. Gong, and J. Zhang, "Privacy-preserving crowdsensing: Privacy valuation, network effect, and profit maximization," in *Proc. IEEE Glob. Commun. Conf. (GLOBECOM)*, 2016, pp. 1–6.
[39] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 2, pp. 486–501, Feb. 2017.
[40] S. Zhong *et al.*, "Connecting human to cyber-world: Security and privacy issues in mobile crowdsourcing networks," in *Proc. Security Privacy Next Gener. Wireless Netw.*, 2019, pp. 65–100.
[41] C. Zhu *et al.*, "Toward big data in green city," *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 14–18, Nov. 2017.
[42] R. Zhu, Z. Li, F. Wu, K. G. Shin, and G. Chen, "Differentially private spectrum auction with approximate revenue maximization," in *Proc. 15th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2014, pp. 185–194.

Guoju Gao received the B.S. degree in information security from the University of Science and Technology of Beijing, Beijing, China, in 2014. He is currently pursuing the Ph.D. degree in computer science and technology with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China.

His current research interests include mobile cloud computing, mobile crowdsourcing, privacy preservation, and incentive mechanism.

Mingjun Xiao (M'13) received the Ph.D. degree in computer application technology from the University of Science and Technology of China (USTC), Hefei, China, in 2004.

He is an Associate Professor with the School of Computer Science and Technology, USTC. He has published more over 60 papers in referred journals and conferences, including in the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON COMPUTERS, INFOCOM, and ICNP. His current research interests include mobile crowdsensing, mobile social networks, vehicular ad hoc networks, mobile cloud computing, auction theory, data security, and privacy.

Dr. Xiao served as the TPC Member of INFOCOM'19, ICDCS'19, INFOCOM'18, ICDCS'15, and Mobihoc'14. He is on the reviewer board of several top journals, such as the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE/ACM TRANSACTIONS ON NETWORKING, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, IEEE TRANSACTIONS ON COMPUTERS, the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and the IEEE TRANSACTIONS ON CLOUD COMPUTING.

Jie Wu (M'90–SM'93–F'09) received the Ph.D. degree in computer engineering from Florida Atlantic University, Boca Raton, Florida, USA, in 1989.

He was a Program Director of the National Science Foundation and a Distinguished Professor at Florida Atlantic University. He is currently the Director of the Center for Networked Computing and a Laura H. Carnell Professor with Temple University, Philadelphia, PA, USA, where he also serves as the Director of International Affairs with the College of Science and Technology, where he served as the Chair of the Department of Computer and Information Sciences from 2009 to 2016 and as an Associate Vice Provost for International Affairs from 2015 to 2017 with Temple University. He regularly publishes in scholarly journals, conference proceedings, and books. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications.

Mr. Wu was a recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award. He serves on several editorial boards, including the IEEE TRANSACTIONS ON SERVICES COMPUTING and the *Journal of Parallel and Distributed Computing*. He was the General Co-Chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, IEEE ICPP 2016, and IEEE CNS 2016, as well as the Program Co-Chair for IEEE INFOCOM 2011 and CCF CNCC 2013. He was an IEEE Computer Society Distinguished Visitor, an ACM Distinguished Speaker, and the Chair for IEEE Technical Committee on Distributed Processing. He is a CCF Distinguished Speaker.

Sheng Zhang (M'11) received the B.S. and Ph.D. degrees in computer science and technology from Nanjing University, Nanjing, China, in 2008 and 2014, respectively.

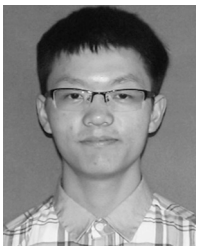
He is an Assistant Professor with Nanjing University, where he is a member of the State Key Laboratory for Novel Software Technology. His publications have appeared in the IEEE TRANSACTIONS ON MOBILE COMPUTING, the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, the IEEE TRANSACTIONS ON COMPUTERS, INFOCOM, ICDCS, and ACM MobiHoc. His current research interests include cloud computing and mobile networks.

Dr. Zhang was a recipient of the Best Paper Runner-Up Award from IEEE MASS 2012.

Liusheng Huang (M'08) received the M.S. degree in computer science from the University of Science and Technology of China, Hefei, China, in 1988.

He is a Professor with the School of Computer Science and Technology, University of Science and Technology of China. He has published 6 books and over 200 papers. His current research interests include delay tolerant networks and Internet of Things.

Prof. Huang serves on the editorial board of many journals.



Guiliang Xiao is currently pursuing the master's degree with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China.

His current research interests include spatial crowdsourcing and blockchain.