Differentially-Private Incentive Mechanism for Crowdsourced Radio Environment Map Construction

Yidan Hu
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
Email: yidanhu@udel.edu

Rui Zhang
Department of Computer and Information Sciences
University of Delaware
Newark, DE 19716
Email: ruizhang@udel.edu

***Abstract*—Database-driven Dynamic Spectrum Sharing (DSS) is a promising technical paradigm for enhancing spectrum efficiency by allowing secondary user to opportunistically access licenced spectrum channels without interfering with primary users' transmissions. In database-driven DSS, a geo-location database administrator (DBA) maintains the spectrum availability in its service region in the form of a radio environment map (REM) and grant or deny secondary users' spectrum access requests based on primary users' activities. Crowdsourcing-based spectrum sensing has great potential in improving the accuracy of the REM at the DBA but requires strong incentives and privacy protection to simulate mobile users' participation. To tackle this challenge, this paper introduces a novel differentially-private reverse auction mechanism for crowdsourcing-based spectrum sensing. The proposed mechanism allows the DBA to select spectrum sensing participants under a budget constraint while offering differential bid privacy, approximate truthfulness, and approximate accuracy maximization. Extensive simulation studies using a real spectrum measurement dataset confirm the efficacy and efficiency of the proposed mechanism.**

I. Introduction

Database-driven Dynamic Spectrum Sharing (DSS) [1], [2] is a promising technical paradigm to meet the ever-growing spectrum demand by allowing secondary user to opportunistically access licensed channels without interfering with primary users' transmissions. In database-driven DSS, a geo-location database administrator (DBA) maintains the spectrum availability in its service region in the form of a radio environment map (REM) [3], [4], where the primary user's received signal strength (RSS) at every location of interest is either directly measured or estimated using proper statistical spatial interpolation techniques. A secondary user can inquire the DBA about a permission to access a licensed spectrum band at the desired location if no primary user is using it.

Constructing and maintaining an accurate REM requires periodically collecting a large number of spectrum measurements over the DBA's service region. A widely advocated approach is for the DBA to deploy only a few dedicated spectrum sensors at selected locations and outsource most spectrum-sensing tasks to pervasive mobile users [5], [6]. The feasibility of this approach is rooted in the deep penetration of mobile devices into people's everyday life. In addition, future mobile devices are widely expected to be capable of spectrum sensing via either internal spectrum sensors or external ones acquired from the DBA [7], [8].

Crowdsourcing-based REM construction requires sound incentive mechanisms to stimulate crowdsourcing workers' participation. In particular, performing spectrum sensing incurs non-trivial effort to crowdsourcing workers, such as their time and device battery. Without strong incentives, potential workers may be reluctant to participate in crowdsourcing-based spectrum sensing. A common approach for providing incentives in mobile crowdsourcing systems is to use reverse auction [9], where crowdsourcing workers sell their services by submitting their bids to the DBA, which in turn selects a subset of bidders as winners and offers payments based on their bids. Reverse auction has been widely used in many mobile crowdsourcing systems such as [10], [11].

A sound reverse auction mechanism for crowdsourcing-based REM construction needs to satisfy three critical requirements. First, crowdsourcing workers are selfish in reality and may lie about their costs if doing so can increase their utilities. This requires the reverse auction mechanism to be *truthful*, which means that bidding the true sensing cost is the optimal strategy for mobile crowdsourcing workers. Second, mobile crowdsourcing workers' bids may reveal their personal information, such as their locations [12], [13] and opportunity costs. While the DBA is commonly assumed to be trusted, curious workers could infer other workers' bids from the change in the payment profiles by submitting different bids for the same sensing task in different rounds [14]. It is thus necessary to protect crowdsourcing workers' bid privacy against other curious workers. Finally, reverse auction involves the selection of a set of winners, which needs to ensure the accuracy of the resulting REM. However, the optimal selection of winners to maximize REM accuracy is an NP-hard problem even without considering the first two requirements. Despite the large body of work on privacy-preserving incentive mechanisms for mobile crowdsourcing systems [15]–[19], none of them satisfy the above three requirements. This situation calls for sound privacy-preserving incentive mechanisms to stimulate crowdsourcing workers' participation while protecting their bid privacy and ensuring high REM accuracy.

In this paper, we tackle this challenge by introducing DPS, a novel differentially-private reverse auction mechanism which can simultaneously ensure the bid privacy for crowdsourcing workers and the accuracy of the constructed REM. In DPS, every crowdsourcing worker submits a bid for performing spectrum sensing at his current location. Serving as the auctioneer, the DBA selects a subset of workers as winners

based on the received bids and determines the payment to the winners. The key ingredient of DPS is a greedy algorithm for selecting a candidate winner set with guaranteed REM accuracy with respect to every possible payment price and choosing the final winner set with corresponding payment price using the exponential mechanism to ensure differential privacy for individual workers. Our main contributions in this paper can be summarized as follows.

- To the best of our knowledge, we are the first to study differentially-private mechanism design for crowdsourcing-based REM construction.
- We introduce a novel differentially-private reverse auction mechanism that can simultaneously provide differential privacy to crowdsourcing workers' bids, approximate truthfulness, and guaranteed REM accuracy at the DBA.
- We thoroughly evaluate the proposed mechanism via a combination of theoretical analysis and detailed simulations studies using real spectrum measurement data, which confirm the efficacy and efficiency of the proposed mechanism.

II. Preliminaries

In this section, we introduce the background on statistical spatial interpolation and Ordinary Kriging [20], the system and adversarial models, crowdsourcing-based REM construction, the auction model, and our design objectives.

A. Background on Statistical Spatial Interpolation and Ordinary Kriging

Kriging [20] refers to a class of geo-statistical spatial interpolation techniques that are originally developed for mining and have been explored for radio mapping in recent years. Kriging models the received signal strength (RSS) at any location $\mathbf{x}$ as a Gaussian random field

$$Z(\mathbf{x}) = \mu(\mathbf{x}) + \delta(\mathbf{x}),$$

where $\mu(\mathbf{x})$ is the mean RSS at location $\mathbf{x}$ capturing path loss and shadowing, and $\delta(\mathbf{x})$ represents possible sampling error.

In Ordinary Kriging (OK) [20], a particular Kriging technique that has been overwhelmingly used for radio mapping [21]–[25], $Z(\mathbf{x})$ is further assumed to be *intrinsic stationary* in the sense that

$$\begin{aligned} &\mathbb{E}[Z(\mathbf{x})] = \mu(\mathbf{x}) = \mu, \\ &\mathbb{E}[Z(\mathbf{x}_1)^2 - Z(\mathbf{x}_2)^2] = 2\gamma(h), \end{aligned} \tag{1}$$

where μ is an unknown constant, $h = ||\mathbf{x}_1 - \mathbf{x}_2||$ is the distance between the two locations, and $\gamma(\cdot)$ is the *semivariogram* function that models the variance between two locations as a function of their distance.

Under OK, the RSS at an unmeasured location $\mathbf{x}_0$ is estimated from the RSSs at measured locations. Specifically, given a set of spectrum measurements at locations $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, the RSS at location $\mathbf{x}_0$ is estimated as

$$\hat{Z}(\mathbf{x}_0) = \sum_{i=1}^{n} \omega_i Z(\mathbf{x}_i), \tag{2}$$

where $\sum_{i=1}^{n} \omega_i = 1$ are normalized weights. It is easy to see that $\hat{Z}(\mathbf{x}_0)$ is a linear unbiased estimator as $\mathbb{E}[\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0)] = \mathbb{E}[\sum_{i=1}^{n} \omega_i Z(\mathbf{x}_i) - Z(\mathbf{x}_0)] = \sum_{i=1}^{n} \omega_i \mathbb{E}[Z(\mathbf{x}_i)] - \mathbb{E}[Z(\mathbf{x}_0)] = \mu \sum_{i=1}^{n} \omega_i - \mu = 0$.

By minimizing the Mean Squared Error (MSE) $\mathbb{E}[(\hat{Z}(\mathbf{x}_0) - Z(\mathbf{x}_0))^2]$ with respect to $\{\omega_i\}$ under the normalization constraint $\sum_{i=1}^{n} \omega_i = 1$, we can obtain a set of linear equations, commonly referred to as Kriging system.

Solving the Kriging system leads to the optimal coefficients given by

$$\omega^* = (\omega_i^*)_{i \in \mathcal{X}} = \Sigma_{\mathcal{X}\mathcal{X}}^{-1} \Sigma_{\mathcal{X}\mathbf{x}_0}, \tag{3}$$

where $\Sigma_{\mathcal{X}\mathcal{X}}^{-1}$ is the covariance matrix, and $\sum_{\mathcal{X}\mathbf{x}_0}$ is the vector of cross-covariances between every $Z(\mathbf{x}_i)(i \in [1, n])$ and $Z(\mathbf{x}_0)$. Since the estimator is unbiased, the minimized MSE, commonly referred to as *Kriging variance (K-var)*, is given by

$$\sigma^2_{\mathbf{x}_0|\mathcal{X}} = \sigma^2_{\mathbf{x}_0} - \Sigma^T_{\mathcal{X}\mathbf{x}_0} (\Sigma^{-1}_{\mathcal{X}\mathcal{X}}) \Sigma_{\mathcal{X}\mathbf{x}_0},$$

where $\sigma^2_{\mathbf{x}_0}$ is the unknown K-var when $\mathcal{X} = \emptyset$. K-var represents the prediction uncertainty at the unmeasured location and is often used as the estimator design metric. The smaller K-var, the higher accuracy of the estimation, and vice versa.

B. System Model

We consider a DBA which maintains an REM for the spectrum availability in its service area $\mathcal{D} \in \mathbb{R}^2$. The area $\mathcal{D}$ is divided into a number of cells of equal size.

The DBA relies on spectrum sensing to construct and maintain the REM. Specifically, the DBA deploys a small number of static spectrum sensors at strategic locations and outsources the majority of spectrum sensing tasks to mobile crowdsourcing workers. Deploying few static spectrum sensors cannot only guarantee minimum level of service when there are insufficient mobile crowdsourcing workers, e.g., during nighttime, but also facilitate detection of potential false spectrum measurements [6]. Denote by $\mathcal{S}$ the set of dedicated spectrum sensors and $\mathcal{N} = \{1, \ldots, n\}$ the set of crowdsourcing workers. We assume that the locations of dedicated spectrum sensors are known to the DBA. We also assume that each crowdsourcing worker owns a mobile device capable of spectrum sensing and acquiring its current location.

The DBA periodically collects spectrum measurements from static spectrum sensors and selected crowdsourcing workers to update the REM. Assume that the time is divided into epoches. At the beginning of each epoch, the DBA broadcasts a spectrum sensing request to all the potential crowdsourcing workers in $\mathcal{D}$, which includes sensing frequency, sampling rate, etc. On receiving the sensing request, each crowdsourcing worker $i \in \mathcal{N}$ submits a bid b_i along with his location $\mathbf{x}_i$ to the DBA, indicating that he is willing to perform spectrum sensing at location $\mathbf{x}_i$ for a minimal payment of b_i. Once the DBA receives a bid-location profile $(b, \mathcal{X})$ where $b = (b_1, \cdots, b_n)$ and $\mathcal{X} = (\mathbf{x}_1, \cdots, \mathbf{x}_n)$, it selects a winner set $\mathcal{W} \subseteq \mathcal{N}$ and determines the payment p_i for each winner $i \in \mathcal{W}$.

The DBA then informs the winners and collects spectrum measurements from them as well as static spectrum sensors. In particular, each static sensor or winning crowdsourcing worker $i \in \mathcal{S} \bigcup \mathcal{W}$ submits a spectrum measurement $Z(\mathbf{x}_i)$ to the

DBA. On receiving all the measurements $\{Z(\mathbf{x}_i)|i \in \mathcal{S} \bigcup \mathcal{W}\}$, the DBA estimates the RSS at the center of every cell using Eq. (2) whereby to produce the updated REM.

C. The Objective Function at the DBA

The DBA's primary objective is to choose the set of winners $\mathcal{W}$ with total payment under the budget constraint while minimizing the average K-var of the produced REM over its service region.

We adopt a similar objective function from [24], where the DBA chooses winners partially based on the predicted contribution of additional measurements submitted at the winners' locations. Specifically, under the optimal weights given in Eq.(3), the K-var at an unmeasured location $\mathbf{x} \in \mathcal{D}$ after taking measurements from deployed dedicated sensors $\mathcal{S}$ at locations $\mathcal{X}_\mathcal{S} = \{\mathbf{x}_i|i \in \mathcal{S}\}$ is given by [24].

$$\sigma^2_{\mathbf{x}|\mathcal{X}_\mathcal{S}} = \sigma^2_\mathbf{x} - \Sigma^T_{\mathcal{X}_\mathcal{S}\mathbf{x}} \Sigma^{-1}_{\mathcal{X}_\mathcal{S}\mathcal{X}_\mathcal{S}} \Sigma_{\mathcal{X}_\mathcal{S}\mathbf{x}}, \tag{4}$$

where $\sigma^2_\mathbf{x}$ is the unknown variance at location $\mathbf{x}$, $\Sigma_{\mathcal{X}_\mathcal{S}\mathcal{X}_\mathcal{S}}$ is the covariance matrix of all measurements from dedicated sensors, and $\Sigma_{\mathcal{X}_\mathcal{S}\mathbf{x}}$ is the vector of cross-covariances between $\{Z(\mathbf{x}_i)|i \in \mathcal{S}\}$ and $Z(\mathbf{x})$.

Given a winner set $\mathcal{W}$, the DBA will collect additional spectrum measurements from locations $\mathcal{X}_\mathcal{W} = \{\mathbf{x}_i|i \in \mathcal{W}\}$. Combining spectrum measurements from $\mathcal{S}$ and $\mathcal{W}$, the Kriging variance at an unmeasured location $\mathbf{x} \in \mathcal{D}$ is given by.

$$\sigma^2_{\mathbf{x}|\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}} = \sigma^2_\mathbf{x} - \Sigma^T_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathbf{x}} \Sigma^{-1}_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}} \Sigma_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathbf{x}}. \tag{5}$$

Subtracting Eq. (5) from Eq. (4), we can obtain the predicted Kriging variance reduction at location $\mathbf{x}$ caused by additional measurements from $\mathcal{W}$ as

$$\begin{aligned} \triangle\sigma^2_\mathbf{x}(\mathcal{W}) = {} & \Sigma^T_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathbf{x}} \Sigma^{-1}_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}} \Sigma_{\mathcal{X}_{\mathcal{S}\cup\mathcal{W}}\mathbf{x}} \\ & - \Sigma^T_{\mathcal{X}_\mathcal{S}\mathbf{x}} \Sigma^{-1}_{\mathcal{X}_\mathcal{S}\mathcal{X}_\mathcal{S}} \Sigma_{\mathcal{X}_\mathcal{S}\mathbf{x}}. \end{aligned} \tag{6}$$

Now consider the whole service region $\mathcal{D}$. The average reduction of Kriging variance caused by the measurements submitted by winner set $\mathcal{W}$ is given by

$$f(\mathcal{W}) = \frac{1}{|\mathcal{D}|} \sum_{\mathbf{x}\in\mathcal{D}} \triangle\sigma^2_\mathbf{x}(\mathcal{W}). \tag{7}$$

Assume that the DBA has a budget B for payment to the winners for each epoch. The DBA intends to find a set of winners $\mathcal{W}$ along with payment profile $\{p_i|i \in \mathcal{W}\}$ under the budget constraint that maximizes the average reduction of Kriging variance in the service region $\mathcal{D}$, which can be formulated as the following optimization problem.

$$\begin{aligned} \text{Maximize} \quad & f(\mathcal{W}) \\ \text{subject to} \quad & \sum_{i\in\mathcal{W}} p_i \le B, \\ & \mathcal{W} \subseteq \mathcal{N}. \end{aligned} \tag{8}$$

The above optimization problem is NP hard. In particular, let us temporally ignore the payment profile and budget constraints and assume that the DBA can choose a fixed number of winners. We can see that even this simplified version of the problem is a special case of subset selection problem, which is NP hard in general because of the non-linear nature of objective function $f(\mathcal{W})$.

D. Other Design Objectives

In addition to budget feasibility and maximizing the average K-var reduction in $\mathcal{D}$, we also intend to design our incentive mechanism to satisfy the following objectives.

Approximate Truthfulness. A selfish crowdsourcing worker may submit bid different from his true valuations of sensing cost if doing so could increase his utility. Assume that each worker i has a true valuation v_i for the cost of performing spectrum sensing at location $\mathbf{x}_i$, which might be different from his bid b_i. The worker i's utility is then given by

$$u_i = \begin{cases} p_i - v_i, & \text{if } i \in \mathcal{W}, \\ 0, & \text{otherwise}, \end{cases} \tag{9}$$

where p_i is the payment worker i receives from the DBA if he is selected as a winner.

As a result, we aim to ensure that every crowdsourcing worker's optimal strategy is to bid his cost truthfully. Exact truthfulness, however, is usually difficult to achieve without losing other desirable properties. Instead, we aim to achieve γ-*truthfulness* such that no crowdsourcing worker can gain more than γ utility by bidding untruthfully.

Definition 1. *(γ-truthful). An auction mechanism is γ-truthful in expectation if and only if for any bid $b_i \neq v_i$ and any bid profile of other workers b_{-i},*

$$\mathbb{E}[u(v_i, b_{-i})] \ge \mathbb{E}[u(b_i, b_{-i})] - \gamma\,. \tag{10}$$

where γ is a small positive constant.

Differential privacy. We also intend to protect crowdsourcing workers' biding privacy. While every worker's bid is known to the DBA and kept private from other workers, a curious worker could still infer other workers' bids by submitting different bids in different rounds of auction. In particular, since the change in a single bid may result in significant change in the selected winner set and the payment profile, a curious worker may infer other workers' bids from the change in the payment he receives from the different payments she receives in different rounds. Differential privacy [9], [26] is a powerful technique to protect bid privacy against such differential attacks. The key idea is that given two neighboring input datasets, a differentially-private mechanism behaves approximately the same on both datasets, such that the presence or absence of a single element would not cause any major change in the output. The formal definition of differential privacy is given as follows.

Definition 2. ***(Differential privacy [9], [26]).*** *Let $M(\cdot)$ be a function that maps an input bid profile b to a payment profile $p \in P$. Mechanism $M(\cdot)$ is ϵ-differentially private if and only if for any set of payment profiles $R \subseteq P$ and any two bid profiles b and b' that differ in only one bid, we have*

$$\mathsf{Pr}[M(b) \in R] \ge \exp(\epsilon)\mathsf{Pr}[M(b') \in R]\,. \tag{11}$$

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

The exponential mechanism [9] is a classical tool to facilitate mechanism design via differential privacy. The key idea is to map a pair of input dataset A and candidate outcome o to a real valued "quality score" $q(A, o)$, where higher score indicates better performance of the outcome. Given the output space $\mathcal{O}$, a score function $q()$, and the privacy budget ϵ, the exponential mechanism chooses the outcome $o \in \mathcal{O}$ with probability proportional to $\epsilon q(A, o)$.

Theorem 1. *[9] The exponential mechanism gives $2\epsilon\triangle$ differential privacy.*

Here $\triangle$ is the *global sensitivity* of $\epsilon q(A, o)$ that captures the largest change in the quality score by a single change of the input in A.

Computation Efficiency. The selection of winner set and corresponding payment price should be computed in polynomial time.

Individual rationality. Our last design objective is individual rationality, which ensures that every crowdsourcing worker's utility is non-negative, i.e., $u_i \geq 0$ for all $i \in \mathcal{N}$, if he bids truthfully. The property is desired to stimulate mobile users's participation in any mobile crowdsourcing systems.

III. The DPS Design

In this section, we first give an overview of DPS and then detail its design.

A. Overview

DPS is designed by integrating a number of ideas. First, inspired by [19], [27], we adopt the single-price mechanism in which the DBA pays every winner the same amount of payment. It has been proved in [28] that the optimal single-price payment mechanism is within a constant factor of any differentiated payment mechanism. Second, under the single-price payment mechanism, we further design a greedy algorithm for selecting winners with guaranteed approximation ratio. Specifically, for any fixed payment price p, the maximum number of workers that the DBA can select is $\lfloor B/p \rfloor$. Any worker whose bid not higher than p can be chosen as a winner without violating the individual rationality. The winner selection problem under the single payment price p is then converted into the special case of subset selection problem which can be solved by greedy algorithm with guaranteed approximation ratio. Third, we choose final winner set and payment price using the exponential mechanism to ensure differential privacy. In particular, for each possible payment price, we can find a corresponding winner set and calculate the predicted average Kriging variance reduction. Given a set of possible payment prices, we then choose the final winner set and payment price using the exponential mechanism.

B. Detailed Design

We now detail the process of winner selection and payment price determination.

On receiving the bid-location profile $(b, \mathcal{X})$, the DBA first finds a set of feasible payment prices. Without loss of generality, we assume that the possible payment to individual worker forms a finite set $P = \{p_{\min}, \ldots, p_{\max}\}$, where the lowest and highest payment prices are $p_{\min}$ and $p_{\max}$, respectively. Let $b_{\min}$ and $b_{\max}$ be the lowest and highest bids in b, respectively. We say a price $p_k \in P$ is feasible if and only if there is at least one crowdsourcing worker with biding price no higher than p_k. The maximum number of winners is constrained by the budget B. In particular, given budget B and payment price p_k, the number of winners is at most $\lfloor B/p_k \rfloor$.

Second, for each feasible payment price $p_k \in P$, the DBA finds a winner set $\mathcal{W}_k$ using a greedy algorithm. The greedy algorithm explores the fact that the objective function $f(\cdot)$ in Eq. (13) is *submodular*, *non-negative*, and *monotone* [24]. Specifically, it is easy to see that the $f(\cdot)$ is non-negative as the K-var reduction is always positive for any non-empty winner set. Moreover, a set function $f : 2^{\mathcal{C}} \to \mathbb{R}$ is submodular if and only if $f(\mathcal{A} \bigcup \{x\}) - f(\mathcal{A}) \geq f(\mathcal{B} \bigcup \{x\}) - f(\mathcal{B})$ for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$ and $x \in \mathcal{C} \setminus \mathcal{B}$. Submodularity captures the diminishing returns behavior of f: adding a new element to the input set always results in the increase in f, and the amount of increase reduces as the number of existing elements increases. Finally, $f(\cdot)$ is monotone if and only if $f(\mathcal{A}) \leq f(\mathcal{B})$ for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$. A widely known result [29] is that for any function that is simultaneously submodular, monotone, and non-negative, a greedy algorithm that chooses the local optimal element at each step can find a solution with guaranteed approximation ratio of $1-1/e$, and no polynomial-time algorithm can achieve a better guarantee unless $P = NP$.

We now detail the greedy algorithm for winner selection for each payment price. Consider payment price p_k as an example, let $\mathcal{N}_k = \{i | b_i \leq p_k\}$ be the set of workers whose bids are not higher than p_k. The DBA maintains a winner set $\mathcal{W}_k$, a set of candidate workers $\mathcal{C}_k$, where $\mathcal{W}_k = \emptyset$ and $\mathcal{C}_k = \mathcal{N}_k$ initially. The winner set is selected in $n_k = \lfloor B/p_k \rfloor$ iterations. In each iteration, the DBA finds worker j from $\mathcal{C}_k$ with

$$j = \arg\max_{j \in \mathcal{C}_k} f(\mathcal{W}_k \bigcup \{j\}) - f(\mathcal{W}_k).$$

In other words, the measurement from winner j is expected to give the maximum K-var reduction among all candidate workers. The DBA then moves worker j from candidate set to the winner set, i.e., $\mathcal{W}_k = \mathcal{W}_k \bigcup \{j\}$ and $\mathcal{C}_k = \mathcal{C}_k \setminus \{j\}$. The algorithm terminates after n_k iterations or $\mathcal{C}_k$ is empty, whichever happens the first.

After computing all possible winner sets $\{\mathcal{W}_k | p_k \in P\}$ using the greedy algorithm, the DBA chooses the final winner set and corresponding payment price using the exponential mechanism to guarantee differential privacy for workers' bids. As discussed in Section II-D, applying the exponential mechanism requires a score function along with its global sensitivity. Here we choose the objective function $f(\cdot)$ as the score function, whose global sensitivity is the maximum change that can be caused by the change in a single bid. In particular, let us represent the greedy algorithm as a function $g(\cdot)$ that takes a bid profile b, a budget B, and a possible payment price p_k as input and outputs a winner set $\mathcal{W}_k$. The function $f \circ g$, i.e., the composition of functions f and g, then maps a bid profile (along with a budget and a payment price) into corresponding K-var reduction. Denote by Δf the *global sensitivity* of $f \circ g$, which we will derive in Section III-C. Given all winner sets $\{\mathcal{W}_k | p_k \in P\}$, the DBA first calculates

the probability distribution

$$\mathsf{Pr}[p = p_k] = \frac{\exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}$$

for all $p_k \in P$, where ϵ is the privacy budget.

The DBA finally chooses the final payment price p_k and corresponding winner set $\mathcal{W}_k$ according the computed probability distribution.

C. Global Sensitivity Δf

We now estimate Δf, the global sensitivity of function $f \circ g$. Directly estimating the global sensitivity of $f(\cdot)$ is unfortunately difficult due to the unpredictable behavior of the greedy algorithm. Instead, we seek to derive an upper bound of Δf that suffices to provide differential privacy guarantee.

Theorem 2. *Let b and b' be any two bid profiles that differ in a single bid. For any payment price $p_k \in P$, let $\mathcal{W}_k$ and $\mathcal{W}'_k$ be the winner sets chosen by the greedy algorithm based on b and b', respectively. We have*

$$|f(\mathcal{W}_k) - f(\mathcal{W}'_k)| \leq (\lfloor B/p_{\min} \rfloor / e + 1)\varphi\,, \tag{12}$$

where $\varphi = \max_{i \in \mathcal{N}} f(\{i\})$.

Proof: Let $\mathcal{C}_k$ and $\mathcal{C}'_k$ be the candidate sets for any payment price $p_k \in P$ under bid profiles b and b', respectively. Since b and b' differ in a single bid, $\mathcal{C}_k$ and $\mathcal{C}'_k$ differ in at most one element. Without loss of generality, suppose that $\mathcal{C}_k = \mathcal{C}'_k \bigcup \{j\}$, e.g., worker j is excluded from $\mathcal{C}'_k$ because $b_j \leq p_k < b'_j$. Now consider the following subset selection problem.

$$\begin{aligned} &\text{Maximize} && f(\mathcal{W}) \\ &\text{subject to} && \mathcal{W} \subseteq \mathcal{C}_k, |\mathcal{W}| = n_k, \end{aligned} \tag{13}$$

where $n_k = \lfloor B/p_k \rfloor$ is the number of winners chosen by the greedy algorithm.

Let $\mathcal{W}_{\text{opt},k}$ and $\mathcal{W}'_{\text{opt},k}$ be the optimal winner sets chosen from $\mathcal{C}_k$ and $\mathcal{C}'_k$, respectively. Also let $\mathcal{W}_k$ and $\mathcal{W}'_k$ be the winner sets chosen from $\mathcal{C}_k$ and $\mathcal{C}'_k$ by the greedy algorithm, respectively. Since $\mathcal{C}_k \supset \mathcal{C}'_k$, we have $\mathcal{W}'_{\text{opt},k} \subset \mathcal{C}_k$, and therefore $f(\mathcal{W}_{\text{opt},k}) \geq f(\mathcal{W}'_{\text{opt},k})$.

Since function $f(\cdot)$ is non-negative, monotone, and submodular, the greedy algorithm can produce a solution within $(1 - \frac{1}{e})$ of the optimal solution. We therefore have $(1 - \frac{1}{e})f(\mathcal{W}_{\text{opt},k}) \leq f(\mathcal{W}_k) \leq f(\mathcal{W}_{\text{opt},k})$ and $(1 - \frac{1}{e})f(\mathcal{W}'_{\text{opt},k}) \leq f(\mathcal{W}'_k) \leq f(\mathcal{W}'_{\text{opt},k})$. It follows that

$$\begin{aligned} |f(\mathcal{W}_k) - f(\mathcal{W}'_k)| &\leq \max\{f(\mathcal{W}'_{\text{opt},k}) - \left(1 - \frac{1}{e}\right) f(\mathcal{W}_{\text{opt},k}), \\ &\qquad f(\mathcal{W}_{\text{opt},k}) - \left(1 - \frac{1}{e}\right) f(\mathcal{W}'_{\text{opt},k})\} \\ &= f(\mathcal{W}_{\text{opt},k}) - \left(1 - \frac{1}{e}\right) f(\mathcal{W}'_{\text{opt},k}), \end{aligned}$$

where the last equation holds because $f(\mathcal{W}'_{\text{opt},k}) \leq f(\mathcal{W}_{\text{opt},k})$.

Let $\varphi = \max_{i \in \mathcal{N}} f(\{i\})$ be the maximal K-var reduction caused by a single worker among all workers. Since $f(\cdot)$ is submodular, we have

$$\begin{aligned} f(\mathcal{W}_{\text{opt},k}) &\leq f(\mathcal{W}'_{\text{opt},k} \bigcup \{j\}) \\ &\leq f(\mathcal{W}'_{\text{opt},k}) + f(\{j\}) \leq f(\mathcal{W}'_{\text{opt},k}) + \varphi. \end{aligned}$$

Since $f(\mathcal{W}_{\text{opt},k}) \leq n_k \varphi$ and $n_k \leq \lfloor B/p_{\min} \rfloor$, it follows that

$$\begin{aligned} |f(\mathcal{W}_k) - f(\mathcal{W}'_k)| &\leq f(\mathcal{W}'_{\text{opt},k}) + \varphi - \left(1 - \frac{1}{e}\right) f(\mathcal{W}'_{\text{opt},k}) \\ &= \frac{1}{e} f(\mathcal{W}'_{\text{opt},k}) + \varphi \leq \left(\frac{n_k}{e} + 1\right)\varphi \\ &\leq (\lfloor B/p_{\min} \rfloor / e + 1)\varphi. \end{aligned}$$

■

IV. Theoretical Analysis

We first have the following theorem regarding DPS's differential privacy guarantee.

Theorem 3. *The DPS auction mechanism is ϵ-differentially private.*

Proof: Let b and b' be two bid profiles that differ in only one worker's bid. For any payment price $p_k \in P$, let $\mathcal{W}_k$ and $\mathcal{W}'_k$ be the winner sets chosen by the greedy algorithm based on b and b', respectively. We have

$$\begin{aligned} \frac{\mathsf{Pr}[M(b) = p_k]}{\mathsf{Pr}[M(b') = p_k]} &= \frac{\frac{\exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}}{\frac{\exp\left(\frac{\epsilon f(\mathcal{W}'_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}'_k)}{2\Delta f}\right)}} \\ &= \frac{\exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\exp\left(\frac{\epsilon f(\mathcal{W}'_k)}{2\Delta f}\right)} \cdot \frac{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}'_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)} \\ &= \exp\left(\frac{\epsilon(f(\mathcal{W}_k) - f(\mathcal{W}'_k))}{2\Delta f}\right) \cdot \frac{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}'_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)} \\ &\leq \exp\left(\frac{\epsilon \Delta f}{2\Delta f}\right) \cdot \frac{\sum_{p_k \in P} \exp\left(\frac{\epsilon(f(\mathcal{W}_k) + \Delta f)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)} \\ &= \exp\left(\frac{\epsilon}{2}\right) \cdot \frac{\exp\left(\frac{\epsilon \Delta f}{2\Delta f}\right) \sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_k \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)} \\ &= \exp\left(\frac{\epsilon}{2}\right) \cdot \exp\left(\frac{\epsilon}{2}\right) \\ &= \exp(\epsilon)\,. \end{aligned} \tag{14}$$

■

Let $\Delta p = p_{\max} - p_{\min}$. We have the following theorem regarding the truthfulness of the auction mechanism.

Theorem 4. *The DPS auction is $\epsilon \Delta p$-truthful.*

Proof: Consider an arbitrary worker $j \in \mathcal{N}$ whose true valuation of the sensing cost is u_j. Let b and b' be two bid profiles that differ in only worker j's bid, e.g., j bids u_j and $b_j \neq u_j$ in b and b', respectively. Similar to the proof of

Theorem 3, for any $p_k \in P$, we have $\mathsf{Pr}[M(b) = p_k] \geq e^{-\epsilon} \cdot \mathsf{Pr}[M(b') = p_k]$. It follows that

$$\begin{aligned}
&\mathbb{E}_{p_k \sim M(b)}[u_i(p_k)] - \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&= \sum_{p_k \in P} u_i(p_k)\mathsf{Pr}[M(b) = p_k] - \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&\geq e^{-\epsilon} \cdot \sum_{p_k \in P} u_i(p_k)\mathsf{Pr}[M(b') = p_k] - \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&= e^{-\epsilon} \cdot \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] - \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&\geq (1-\epsilon) \cdot \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] - \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&= -\epsilon \cdot \mathbb{E}_{p_k \sim M(b')}[u_i(p_k)] \\
&= -\epsilon \Delta p,
\end{aligned} \tag{15}$$

where the last inequality holds because $u_i(p_k) \leq p_{\max} - p_{\min} = \Delta p$. We therefore conclude that DPS is $\epsilon\Delta p$-truthful. ■

Finally, we have the following theorem regarding the quality of the REM produced by the auction mechanism.

Theorem 5. *Let $\mathcal{W}_{opt}$ be the optimal winner set among all possible winner sets $\{\mathcal{W}_p | p \in P\}$. Assume that DPS selects a winner set $\mathcal{W}_k$ with payment price p_k. The expected average K-var reduction given by $\mathcal{W}_k$ and the maximum average K-var reduction given by $f(\mathcal{W}_{opt})$ satisfies that*

$$\mathbb{E}_{p_k \in P}[f(\mathcal{W}_k)] \geq f(\mathcal{W}_{opt}) - \ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{opt})}{2\Delta f}\right) \times \left(\frac{6\Delta f}{\epsilon}\right).$$

Proof: We start by defining the following four sets for any constant $t > 0$, including $\mathcal{B}_t = \{p_k | f(\mathcal{W}_k) > f(\mathcal{W}_{\text{opt}}) - t\}, \bar{\mathcal{B}}_t = \{p_k | f(\mathcal{W}_k) \leq f(\mathcal{W}_{\text{opt}}) - t\}, \mathcal{B}_{2t} = \{p_k | f(\mathcal{W}_k) > f(\mathcal{W}_{\text{opt}}) - 2t\}$, and $\bar{\mathcal{B}}_{2t} = \{p_k | f(\mathcal{W}_k) \leq f(\mathcal{W}_{\text{opt}}) - 2t\}$.

Since $\mathsf{Pr}[p_k \in \mathcal{B}_t] \leq 1$, we have

$$\begin{aligned}
\mathsf{Pr}[p_k \in \bar{\mathcal{B}}_{2t}] &\leq \frac{\mathsf{Pr}[p_k \in \bar{\mathcal{B}}_{2t}]}{\mathsf{Pr}[p_k \in \mathcal{B}_t]} \\
&= \frac{\sum_{p_k \in \bar{\mathcal{B}}_{2t}} \frac{\exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_i \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_i)}{2\Delta f}\right)}}{\sum_{p_k \in \mathcal{B}_t} \frac{\exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_i \in P} \exp\left(\frac{\epsilon f(\mathcal{W}_i)}{2\Delta f}\right)}} \\
&= \frac{\sum_{p_k \in \bar{\mathcal{B}}_{2t}} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)}{\sum_{p_k \in \mathcal{B}_t} \exp\left(\frac{\epsilon f(\mathcal{W}_k)}{2\Delta f}\right)} \\
&< \frac{\sum_{p_k \in \bar{\mathcal{B}}_{2t}} \exp\left(\frac{\epsilon f(\mathcal{W}_{\text{opt}}) - 2t)}{2\Delta f}\right)}{\sum_{p_k \in \mathcal{B}_t} \exp\left(\frac{\epsilon f(\mathcal{W}_{\text{opt}}) - t)}{2\Delta f}\right)} \\
&< \frac{|\bar{\mathcal{B}}_{2t}| \exp\left(\frac{\epsilon (f(\mathcal{W}_{\text{opt}}) - 2t)}{2\Delta f}\right)}{|\mathcal{B}_t| \exp\left(\frac{\epsilon (f(\mathcal{W}_{\text{opt}}) - t)}{2\Delta f}\right)} \\
&= \frac{|\bar{\mathcal{B}}_{2t}|}{|\mathcal{B}_t|} \exp\left(\frac{-\epsilon t}{2\Delta f}\right).
\end{aligned}$$

Since $\mathsf{Pr}[p_k \in \bar{\mathcal{B}}_{2t}] + \mathsf{Pr}[p_k \in \mathcal{B}_{2t}] = 1$, it follows that

$$\begin{aligned}
\mathsf{Pr}[p_k \in \mathcal{B}_{2t}] &= 1 - \mathsf{Pr}[p_k \in \bar{\mathcal{B}}_{2t}] \\
&> 1 - \frac{|\bar{\mathcal{B}}_{2t}|}{|\mathcal{B}_t|} \exp\left(\frac{-\epsilon t}{2\Delta f}\right).
\end{aligned}$$

We can estimate the $\mathbb{E}_{p_k \in P}[f(\mathcal{W}_k)]$ as

$$\begin{aligned}
\mathbb{E}_{p_k \in P}[f(\mathcal{W}_k)] &= \sum_{p_k \in P} f(\mathcal{W}_k)\mathsf{Pr}[p = p_k] \\
&\geq \sum_{p_k \in \mathcal{B}_{2t}} f(\mathcal{W}_k)\mathsf{Pr}[p = p_k] \\
&\geq (f(\mathcal{W}_{\text{opt}}) - 2t)\mathsf{Pr}[p_k \in \mathcal{B}_{2t}] \\
&\geq (f(\mathcal{W}_{\text{opt}}) - 2t)\left(1 - \frac{|\bar{\mathcal{B}}_{2t}|}{|\mathcal{B}_t|}\exp\left(\frac{-\epsilon t}{2\Delta f}\right)\right) \\
&\geq (f(\mathcal{W}_{\text{opt}}) - 2t)\left(1 - |P|\exp\left(\frac{-\epsilon t}{2\Delta f}\right)\right),
\end{aligned} \tag{16}$$

where the last inequality holds as $|\bar{\mathcal{B}}_{2t}| \leq |P|$ and $|\mathcal{B}_t| \geq 1$.

For any t satisfying

$$t \geq \ln\left(\frac{f(\mathcal{W}_{\text{opt}})|P|}{t}\right) \times \left(\frac{2\Delta f}{\epsilon}\right), \tag{17}$$

we have

$$\begin{aligned}
\exp\left(\frac{-\epsilon t}{2\Delta f}\right) &\leq \exp\left(\frac{-\epsilon\left(\ln\left(\frac{f(\mathcal{W}_{\text{opt}})|P|}{t}\right) \times \left(\frac{2\Delta f}{\epsilon}\right)\right)}{2\Delta f}\right) \\
&= \frac{t}{f(\mathcal{W}_{\text{opt}})|P|}.
\end{aligned} \tag{18}$$

Plugging Inequality. (18) into Eq. (16), we get

$$\begin{aligned}
\mathbb{E}_{p_k \in P}[f(\mathcal{W}_k)] &= (f(\mathcal{W}_{\text{opt}}) - 2t)\left(1 - |P|\exp\left(\frac{-\epsilon t}{2\Delta f}\right)\right) \\
&\geq (f(\mathcal{W}_{\text{opt}}) - 2t)\left(1 - |P| \cdot \frac{t}{f(\mathcal{W}_{\text{opt}})|P|}\right) \\
&= f(\mathcal{W}_{\text{opt}}) - 3t + \frac{2t^2}{f(\mathcal{W}_{\text{opt}})} \\
&> f(\mathcal{W}_{\text{opt}}) - 3t,
\end{aligned} \tag{19}$$

if Inequality (17) holds.

We now show that $t = \ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) \times \left(\frac{2\Delta f}{\epsilon}\right)$ satisfies Inequality (17). In particular, since $\ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) > 1$, we have $t > \frac{2\Delta f}{\epsilon}$. In addition, since $\ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) > \ln\left(\frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right)$, we have

$$\begin{aligned}
t &= \ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) \times \left(\frac{2\Delta f}{\epsilon}\right) \\
&\geq \ln\left(|P|f(\mathcal{W}_{\text{opt}}) \cdot \frac{\epsilon}{2\Delta f}\right) \times \left(\frac{2\Delta f}{\epsilon}\right) \\
&> \ln\left(\frac{|P|f(\mathcal{W}_{\text{opt}})}{t}\right) \times \left(\frac{2\Delta f}{\epsilon}\right).
\end{aligned}$$

TABLE I. DEFAULT SIMULATION SETTING

Para.	Val.	Description.
$p_{\min}$	1	The lowest payment price
$p_{\max}$	2	The highest payment price
$b_{\min}$	2	The lowest bid price
$b_{\max}$	2	The highest bid price
$\|P\|$	101	The number of possible payment prices
$\|\mathcal{S}\|$	5	The number of dedicated sensors
ϵ	0.1	Privacy budget
$\|\mathcal{N}\|$	140	The number of crowdsourcing workers
B	30	Budget

Finally, substituting $t = \ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) \times \left(\frac{2\Delta f}{\epsilon}\right)$ into Eq. (19), we obtain

$$\mathbb{E}_{p_k \in P}[f(\mathcal{W}_k)] \geq f(\mathcal{W}_{\text{opt}}) - \ln\left(e + \frac{\epsilon|P|f(\mathcal{W}_{\text{opt}})}{2\Delta f}\right) \times \left(\frac{6\Delta f}{\epsilon}\right).$$

The theorem is therefore proved. ■

We finally have the following theorem with the proof omitted due to space constraints.

Theorem 6. *The DPS auction mechanism achieves budget feasibility and individual rationality and can compute the winner set and payment price in polynomial time.*

V. SIMULATION RESULTS

In this section, we evaluate the performance of DPS via simulation using a real spectrum measurement dataset.

A. Dataset

As in [6], [23], we use the CRAWDAD `cu/wimax` dataset [30] for our simulation studies. The `cu/wimax` dataset was collected at the University of Colorado Boulder (UC) and contains the signal-to-interference-plus-noise ratio (CINR) measurements of five WiMax base stations serving the University of Colorado campus taken on a 100m equilateral triangular lattice. For our purpose, we chose the total 145 measurements for channel 308 and BSID 3674210305.

B. Simulation Settings

We randomly divide the total 145 measurements into a set of 5 measurements as the ones reported by dedicated anchor sensors and a set of the remaining 140 as submitted by mobile crowdsourcing workers. We fit the semivariogram from the total 145 measurements along with their locations. We also assume that the semivariogram of each location is to the DBA. In addition, the bid price of each mobile user is randomly picked among $\{1, 1.01, \ldots, 2\}$. Every point in the following figures is the average of 100 runs, each with a distinct seed. Table I summarizes our default simulation settings unless mentioned otherwise.

Since DPS is the first solution for crowdsourced REM construction, we compare DPS with other two strategies.

- **Baseline differentially private auction (BDPA)**: In BDPA, for each possible price $p_k \in P$, the DBA first computes predicted average K-var reduction $f(\{i\})$ for each worker $i \in \mathcal{N}_k$ and selects winner set $\mathcal{W}_k$ as the $\lfloor B/p_k \rfloor$ workers with the highest average K-var reductions. The final winner set and payment price are chosen using the exponential mechanism as in the DPS. It is easy to verify that BDPA achieves approximate truthfulness and ϵ differential privacy.
- **Optimal single-price auction (OSPA)**: In OSPA, for each possible price $p_k \in P$, the DBA chooses the corresponding winner set $\mathcal{W}_k$ using the greedy algorithm as in DPS and then selects the final winner set with corresponding payment price as the one that gives the maximum average K-var reduction. The K-var reduction achieved by OSPA can be viewed as the upper bound of the DPS.

We use two metrics to evaluate the performance of DPS: *average K-var reduction* and *privacy leakage*. Besides the average K-var reduction defined in Section II, the privacy leakage is defined as follows.

Privacy Leakage. We use the Kullback-Leibler divergence [31] to evaluate the the privacy leakage of DPS. Let b and b' be two bid profiles that differ in a single bid. Denote their payment probability distributions under DPS as $\mathsf{Pr}[M(b)]$ and $\mathsf{Pr}[M(b')]$, respectively. The privacy leakage in terms of the Kullback-Leibler divergence is defined as

$$\begin{aligned}\mathsf{PL} &= KL\big(\mathsf{Pr}[M(b)] \| \mathsf{Pr}[M(b')]\big) \\ &= \sum_{p_k \in P} \mathsf{Pr}[M(b) = p_k] \ln\left(\frac{\mathsf{Pr}[M(b) = p_k]}{\mathsf{Pr}[M(b') = p_k]}\right).\end{aligned}$$

KL divergence indicates the statistical difference between two probability distributions. Generally speaking, the smaller KL, the harder to distinguish the two bid profiles and thus better protection of workers' bid privacy.

C. Simulation Results

1) Impact of Budget B: Fig. 1 compares the K-var reductions under BDPA, OSPA, and DPS with total budget B varying from 10 to 80. As we can see, as the total budget increases, the average K-var reductions of all three mechanisms increase. This is anticipated, as the higher budget, the more winners chosen by the DBA, the higher average K-var reduction, and vice versa. Moreover, the OSPA's K-var reduction is always the highest, which confirms that it is the upper bound of the DPS mechanism. While DPS's average K-var reduction is slightly lower than that of OSPA, it outperforms BDPA by a large margin. These results indicate that DPS can achieve approximate maximal K-var reduction while providing differential bid privacy to crowdsourcing workers.

2) Impact of Privacy Budget ϵ: Fig. 2 compares the K-var reductions of BDPA and DPS varying with privacy budget ϵ, where the K-var reduction of AMNDP is not affected by the change in ϵ and is plotted for reference only. As we can see, the K-var reductions of DPS and BDPA both increase as ϵ increases. The reason is that the larger ϵ, the higher the probability of high-quality winner set and payment price being selected by the exponential mechanism, the higher K-var reduction, and vice versa. Moreover, the variance of K-var

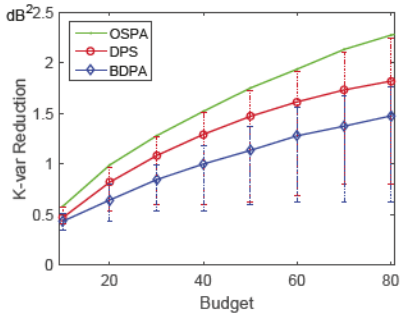

Fig. 1. K-var reduction vs. budget B.

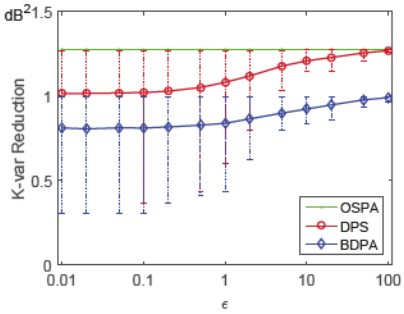

Fig. 2. K-var reduction vs. privacy budget ϵ.

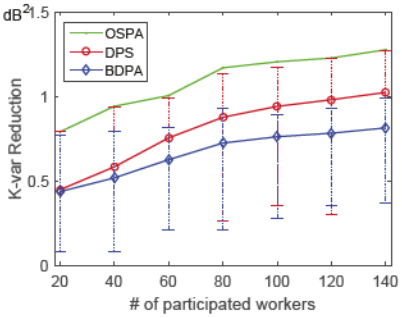

Fig. 3. K-var reduction vs. # of workers.

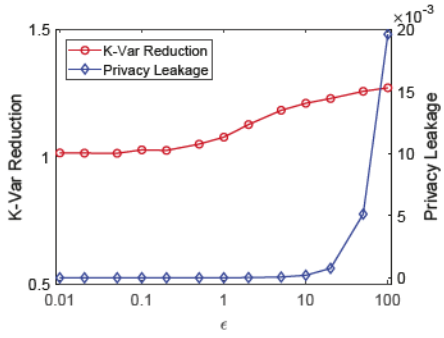

Fig. 4. Privacy leakage vs. ϵ.

reductions of both BDPA and DPS decrease as ϵ increases, which is anticipated. In addition, the K-var of DPS is always higher than that of BDPA by a large margin, which confirms the effectiveness of the greedy algorithm in selecting high-quality winner set.

3) Impact of the Number of Workers: Fig. 3 compares the K-var reductions of BDPA, OSPA, and DPS as the number of participating workers increases from 20 to 140. We can see that the K-var reductions of all three mechanisms increase as the number of participating workers increases, as the DBA can select more winners. Similar to Fig. 1, the OSPA's K-var reduction is always the highest, followed by DPS, and that of BDPA is the lowest. It is worth noting that when the number of participating workers is small, the advantage of DPS over OSPA is small. For example, when the number of participating workers is 20, DPS and OSPA have the same K-var reduction. This is because the DBA can afford to select all the workers as winners in such cases. Finally, the difference between DPS and AMNDP is caused by the exponential mechanism and can be viewed as the cost of providing differential bid privacy.

4) Privacy Leakage: Fig. 4 shows the privacy leakage and K-var reduction under DPS varying with privacy budget ϵ. As we can see, as ϵ increases, the privacy leakage and K-var reduction both increase. This is expected, as the larger ϵ, the higher the probability of high-quality winner set being selected by the exponential mechanism, the higher K-var reduction, and vice versa. At the same time, the higher ϵ, the less privacy protection, and the larger privacy leakage, and vice versa. Generally speaking, the choice of ϵ represents a trade-off between the quality of winner set (i.e., REM's accuracy) and privacy leakage.

VI. Related Work

Differentially-private mechanism design has attracted much attention in recent years. The first differentially-private auction mechanism was introduced in [9] by incorporating the exponential mechanism. Several general differentially-private auction mechanisms with the goal of maximizing social welfare were presented in [15], [16], [32], [33]. In [27], [34], Zhu *et al.* studied differentially-private spectrum auction mechanism with approximate truthfulness and approximate revenue maximization. BidGuard [14] is a differentially private auction mechanism aiming at minimizing social cost. Jin *et al.* [35] proposed a differentially-private incentive mechanism for minimizing total payment while ensuring approximate truthfulness and individual rationality. More recently, Jin *et al.* [36] introduced a differentially-private double auction mechanism for mobile crowdsensing systems with platform revenue maximization. None of these solutions can be applied to crowdsourced REM construction because of their very different objectives.

Privacy-preserving auction mechanism has also been studied for spectrum allocation problem. THEMIS [37] incorporates cryptographic technique into spectrum auction to deal with the seller-side fraudulent actions. Huang *et al.* [18] introduced a truthful and privacy-preserving mechanism to achieve k-anonymity in spectrum auctions. Subsequently, PPS [17] applied homomorphic encryption for maximizing the social efficiency and preserving bid privacy. These solutions rely on expensive cryptographic techniques and do not offer differential privacy guarantee for individual worker's bid.

Truthful incentive mechanism for general mobile crowdsourcing systems has been an active research area. Yang *et al.* [10] introduced truthful incentive mechanisms for mobile crowdsensing systems with the goal of maximizing platform utility. Zhao *et al.* [11] studied an similar problem under the online auction model. TRAC [38] is a truthful reverse auction mechanism for location-aware crowdsensing systems. Moreover, truthful double auction was studied in [39] in the context of crowdsourcing systems involving multiple data requesters. Our work is mostly related to [24], in which Ying *et al.* introduced an incentive mechanism for crowdsourcing-based REM construction with approximate maximization of K-var reduction. However, none of these solutions consider protecting worker's bid privacy.

There are also some work loosely related to our work. Yang *et al.* [40] introduced an truthful auction mechanism to incentivize mobile users to participate in anonymity set

to achieve k-anonymity. A truthful spectrum double auctions mechanism was introduced in [41] for maximizing platform profit. INCEPTION [42] is a truthful reverse auction for crowdsourcing-based data aggregation which incorporates data perturbation mechanism to reduces workers privacy leakage.

VII. Conclusion

In this paper, we have introduced the design and evaluation of DPS, a novel differentially-private reverse auction mechanism for crowdsourced REM construction. We have proved that the proposed auction mechanism achieves approximate truthfulness, differential privacy, and near-optimal REM accuracy. Extensive simulation studies using a real spectrum measurement dataset confirm the efficacy and efficiency of the proposed mechanism.

References

[1] D. Gurney, G. Buchwald, L. Ecklund, S. L. Kuffner, and J. Grosspietsch, "Geo-location database techniques for incumbent protection in the tv white space," in *IEEE DySPAN'08*, Oct 2008, pp. 1–9.

[2] R. Murty, R. Chandra, T. Moscibroda, and P. V. Bahl, "Senseless: A database-driven white spaces network," *IEEE Transactions on Mobile Computing*, vol. 11, no. 2, pp. 189–203, 2012.

[3] Y. Zhao, L. Morales, J. Gaeddert, K. K. Bae, J. S. Um, and J. H. Reed, "Applying radio environment maps to cognitive wireless regional area networks," in *IEEE DySPAN'07*, April 2007, pp. 115–118.

[4] H. B. Yilmaz, T. Tugcu, F. Alagoz, and S. Bayhan, "Radio environment map as enabler for practical cognitive radio networks," *IEEE Communications Magazine*, vol. 51, no. 12, pp. 162–169, December 2013.

[5] R. Zhang, J. Zhang, Y. Zhang, and C. Zhang, "Secure crowdsourcing-based cooperative spectrum sensing," in *IEEE INFOCOM'13*, Apr. 2013.

[6] Y. Hu and R. Zhang, "Secure crowdsourced radio environment map construction," in *IEEE ICNP'17*, Toronto, Canada, Oct. 2017.

[7] R. Calvo-Palomino, D. Pfammatter, D. Giustiniano, and V. Lenders, "A low-cost sensor platform for large-scale wideband spectrum monitoring," in *IPSN'15*, Seattle, Washington, 2015, pp. 396–397.

[8] D. Pfammatter, D. Giustiniano, and V. Lenders, "A software-defined sensor architecture for large-scale wideband spectrum monitoring," in *IPSN'15*, Seattle, Washington, 2015, pp. 71–82.

[9] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *FOCS'07*, Washington, DC, USA, 2007, pp. 94–103.

[10] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *ACM MobiCom'12*, Istanbul, Turkey, 2012, pp. 173–184.

[11] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *IEEE INFOCOM'14*, Toronto, Canada, Apr. 2014.

[12] X. Jin and Y. Zhang, "Privacy-preserving crowdsourced spectrum sensing," in *IEEE INFOCOM'16*, San Francisco, CA, 2016.

[13] T. Wen, Y. Zhu, and T. Liu, "P2: A location privacy-preserving auction mechanism for mobile crowd sensing," in *IEEE GLOBECOM'16*, Washington, DC, USA, Dec. 2016.

[14] J. Lin, D. Yang, M. Li, J. Xu, and G. Xue, "BidGuard: A framework for privacy-preserving crowdsensing incentive mechanisms," in *IEEE CNS'16*, Philadelphia, PA, USA, Apr. 2016.

[15] K. Nissim, R. Smorodinsky, and M. Tennenholtz, "Approximately optimal mechanism design via differential privacy," in *ITCS'12*, Cambridge, MA, 2012, pp. 203–213.

[16] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," in *EC'13*, Philadelphia, PA, 2013, pp. 215–232.

[17] H. Huang, X. Y. Li, Y. e. Sun, H. Xu, and L. Huang, "PPS: Privacy-preserving strategyproof social-efficient spectrum auction mechanisms," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 5, pp. 1393–1404, May 2015.

[18] Q. Huang, Y. Tao, and F. Wu, "SPRING: a strategy-proof and privacy preserving spectrum auction mechanism," in *INFOCOM'13*, April 2013.

[19] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *IEEE ICDCS'16*, Nara, Japan, June 2016.

[20] N. A. Cressie and N. A. Cassie, *Statistics for spatial data*, Wiley-Interscience; 2 edition, July 1993.

[21] A. B. H. Alaya-Feki, S. B. Jemaa, B. Sayrac, P. Houze, and E. Moulines, "Informed spectrum usage in cognitive radio networks: Interference cartography," in *PIMRC'08*, Sept 2008, pp. 1–5.

[22] A. Achtzehn, J. Riihijarvi, G. M. Vargas, M. Petrova, and P. Mahonen, "Improving coverage prediction for primary multi-transmitter networks operating in the tv whitespaces," in *IEEE SECON'12*, June 2012.

[23] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *IEEE DYSPAN'12*, Oct 2012, pp. 422–433.

[24] X. Ying, S. Roy, and R. Poovendran, "Incentivizing crowdsourcing for radio environment mapping with statistical interpolation," in *IEEE DySPAN'15*, Sept 2015, pp. 365–374.

[25] X. Ying, C. W. Kim, and S. Roy, "Revisiting tv coverage estimation with measurement-based statistical interpolation," in *COMSNETS'15*, Jan 2015, pp. 1–8.

[26] C. Dwork, "Differential privacy," in *ICALP'06*, S. Servolo, Venice, Italy, July 2006, pp. 1–12.

[27] R. Zhu, Z. Li, F. Wu, K. Shin, and G. Chen, "Differentially private spectrum auction with approximate revenue maximization," in *ACM MobiHoc'14*, Aug. 2014, pp. 185–194.

[28] S. Rathinakumar and M. K. Marina, "Gavel: Strategy-proof ascending bid auction for dynamic licensed shared access," in *ACM MobiHoc'16*, July 2016.

[29] A. Das and D. Kempe, "Algorithms for subset selection in linear regression," in *STOC'08*, Victoria, British Columbia, Canada, 2008.

[30] M. Ton and C. Phillips, "CRAWDAD dataset cu/wimax (v.2012-06-01)," http://crawdad.org/cu/wimax/20120601, Jun. 2012.

[31] S. Kullback and R. A. Leibler, "On information and sufficiency," *Ann. Math. Statist.*, vol. 22, no. 1, pp. 79–86, 03 1951.

[32] S. Leung and E. Lui, "Bayesian mechanism design with efficiency, privacy, and approximate truthfulness," in *WINE'12*, Berlin, Heidelberg, 2012, pp. 58–71.

[33] D. Xiao, "Is privacy compatible with truthfulness?" in *ITCS'13*, Berkeley, CA, 2013, pp. 67–86.

[34] R. Zhu and K. G. Shin, "Differentially private and strategy-proof spectrum auction with approximate revenue maximization," in *IEEE INFOCOM'15*, Apr. 2015, pp. 918–926.

[35] H.Jin, L.Su, B.Ding, K.Nahrstedt, and N.Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *IEEE ICDCS'16*, Nara, Japan, 2016, pp. 344–353.

[36] W. Jin, M. Li, L. Guo, and L. Yang, "DPDA: A differentially private double auction scheme for mobile crowd sensing," in *IEEE CNS'18*, Beijing, China, May-June 2018.

[37] M. Pan, J. Sun, and Y. Fang, "Purging the back-room dealing: Secure spectrum auction leveraging paillier cryptosystem," *IEEE J. Sel. Areas Commun.* , vol. 29, no. 4, pp. 866–876, April 2011.

[38] Z. Feng, Y. Zhu, Q. Zhang, L. Ni, and A. V. Vasilakos, "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *IEEE INFOCOM'14*, Toronta, Canada, April 2014.

[39] H. Jin, L. Su, and K. Nahrstedt, "CENTURION: incentivizing multi-requester mobile crowd sensing," in *IEEE INFOCOM'17*, Atlanta, GA, May 2017, pp. 1–9.

[40] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k-anonymity location privacy," in *IEEE INFOCOM'13*, Turin, Italy, Apr. 2013.

[41] D. Yang, X. Zhang, and G. Xue, "Promise: A framework for truthful and profit maximizing spectrum double auctions," in *IEEE INFOCOM'14*, Toronto, ON, Canada, May 2014.

[42] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *ACM MobiHoc'16*, July 2016.