Incentivizing Crowdsourcing for Radio Environment Mapping with Statistical Interpolation

Xuhang Ying, Sumit Roy and Radha Poovendran
Department of Electrical Engineering, University of Washington
Email: {xhying, sroy, rp3}@uw.edu

***Abstract*—White Space Networking crucially relies on the active monitoring of spectrum usage (to identify white space opportunities) in both space and time. One way to achieve this is wide-area deployment of spectrum sensors to gather spatio-temporal spectrum data, and use them to construct better Radio Environment Maps (REMs) via suitable statistical interpolation techniques (i.e., Kriging). Cost of such large-scale sensor deployment can be reduced via *crowdsourcing*, i.e., outsourcing the sensing task to mobile users equipped with sensorized high-end client devices (e.g., tablets or smartphones), and success of such crowdsourced sensing presumes some *incentive mechanisms* to attract user participation. In this work, we present an incentivized crowdsourcing system architecture that (periodically) acquires spectrum data from users, so as to optimize the resulting radio environment map (i.e., minimizing the average prediction-error variance) for a given data acquisition budget. First, we introduce an auction-based incentive mechanism that is computationally efficient, individually rational and truthful, and prove that the total payment of the proposed mechanism is a monotonically increasing function of the cardinality of the winner set. Then we propose a budget-feasible version and through extensive simulations, we evaluate the performance of proposed mechanisms for comparison to a baseline to demonstrate its significantly superior performance in crowdsourced radio mapping.**

***Index Terms*—Incentive Mechanism, Crowdsourcing, Radio Environment Mapping, Statistical Interpolation, Kriging.**

I. Introduction

The exponential growth of high-end client mobile devices translates to a proportionate surge in a need for network capacity, and consequently, additional spectrum. However, within today's static spectrum allocation policy regime, an overwhelming portion of the licensed spectrum is already allocated (primarily to government users). A succession of spectrum utilization studies worldwide have demonstrated that spectrum is often grossly underutilized by licensed (primary) users. To improve spectrum utilization, the concept of Dynamic Spectrum Access based on Cognitive Radio (CR) technology has been proposed, which allows unlicensed (secondary) users to opportunistically access, on a temporal basis, locally idle spectrum[1] subject to the no-harmful-interference-to-primary-user constraint as specified by the Federal Communication Commission (FCC) [1].

White Space Networking is fundamentally based on real-time spectrum usage monitoring and the ability to identify white spaces in space and time. Predictive radio propagation modeling - which primarily takes transmitter parameters (e.g., transmission power, location, antenna pattern) - is often a useful starting point to (pre)-compute the received signal strength (RSS) at any receiver location and build a radio map database [1]. Due to its efficiency and scalability, such modeling - widely used in academia - has also been adopted by FCC for unlicensed access in TV bands [1]. However, since empirical models do not count for finer-grained local environment details (e.g., trees or buildings), RSS predictions tend to be locally inaccurate in many circumstances such as urban areas. Moreover, such modeling captures largely spatial variations of spectrum usage, but not temporal variations in cases where the primary transmitters are dynamic.

Clearly, the above implies the need for wide-area *spectrum sensing* via (either static or dynamic) sensors that are able to provide more accurate local spatio-temporal RSS data, which can be used to construct better REMs [2]–[4]. Such spectrum measurement infrastructure samples the (unknown) RSS map at different locations at a time instant and applies model-based statistical interpolation techniques (e.g., Kriging) to estimate RSS values at unmeasured locations. From this perspective, the REM construction is a *spatial sampling* problem, where sample locations can be carefully selected to achieve desired estimation (REM accuracy) performance under a certain sampling budget (e.g., labor and time).

A core question is how and where to deploy sensors over a desired region for REM creation to capture spatial or temporal variations. Without any a-priori information about the RSS field, the obvious baseline is to deploy static sensors uniformly at random (with density subject to cost constraints) over the region of interest. However, scaling such deployments inevitably meets cost limitations. Hence, the preferred solution is to exploit crowdsourcing [5], namely, outsourcing wide-area sensing to spatially distributed users with mobile devices that are outfitted with spectrum sensors. Recently, this approach has gained a growing attention from the research community. As an example, in [6], authors designed low-cost spectrum sensing hardware for crowdsourcing. In [7], security issues resulting from falsification of crowdsourced sensing data were investigated. Recently, crowdsourcing-enabled spectrum profiling was formulated as a sensing task allocation problem [8], so as to maximize the collective utility of the sensing data.

Prior works on crowdsourced spectrum sensing implicitly assume voluntary participation of cognitive or secondary users. However, since users have to contribute their own resources (e.g., computing power, storage and battery) for sensing, they would expect some form of compensation for their partici-

This work was supported in part by NSF AST award 1443923 under the EARS program.

[1]Such Cognitive usage is often termed as White Space Networks as shorthand.

pation. On the other hand, the platform (offeror) may also have a budget that limits the total payment to users. In this paper, we focus on the design of *truthful* and *budget-feasible* incentive mechanisms for crowdsourced radio mapping with the Kriging-based REM estimation approach, given a budget constraint for the offeror. Since the REM estimation is inherently a spatial sampling problem, sample locations need to be chosen in a manner that minimizes the average prediction-error variance. Similar to [9]–[11], we consider the problem of incentivized crowdsourcing for radio mapping within a *reverse auction* framework, where multiple users want to sell their location-specific spectrum data to the platform. Each user has a privately known *cost* and receives a *payment* when selected. We assume that users are *self-interested* and want to maximize their own *utility*. As a result, they tend to take strategic actions and ask for higher prices (called *bids*) than their true costs. Hence, we are interested in designing a truthful mechanism that motivates users to reveal their true costs via their bids and is also budget-feasible.

Our primary contributions are as follows:

- First, we design a crowdsourcing system that periodically acquires spectrum data from users to construct REMs with statistical interpolation under a budget constraint.
- Second, we propose an auction-based budget-free mechanism with a cardinality constraint (i.e., the maximum number of selected users), and show that it is truthful as well as computationally efficient and individually rational. On top of it, we propose a budget-feasible mechanism by translating the budget constraint to the best cardinality constraint using the bisection method.
- Finally, we conduct extensive simulations to evaluate the performance of the proposed mechanism, and compare it against the state-of-art budget-feasible mechanism proposed in [11]. Our results reveal that the proposed mechanism makes full use of the budget, and performs significantly better compared to the baseline, with an improvement of 18%-22% in terms of maximizing average prediction-error variance reduction.

The remaining of this paper is organized as follows. A review of related work is provided in Section II, and preliminaries are given in Section III. In Sections IV and V, we describe our system and present our mechanism, respectively. Simulation results are given in Section VI, while we provide our conclusion is in Section VII.

II. Related Work

In crowdsourcing, users are typically assumed to be self-interested and tend to take strategic actions. Therefore, it is desirable to design a truthful mechanism that motivates users to tell their true costs. In [9], authors proposed an auction-based truthful mechanism for a scenario where sensing tasks have predetermined location tags and values, and users are only allowed to compete for tasks within their own coverage regions. However, values of collected data did not depend on user locations and the budget was not imposed as a hard constraint in addition to truthfulness. In [12] and [13], similar geometric coverage models for tasks and users were considered, which do not fit into our radio mapping scenario. Instead, we consider a general task without location tags and all users are allowed to compete for it. Based on user locations, the spatial sampling approach is taken to select users.

In [10], [11], authors proposed truthful and budget-feasible incentive mechanisms for general submodular monotone functions. Their mechanisms adopt the proportional share allocation rule, and winner selection stops when the bid of next user exceeds a proportional share of its contribution. Although this rule provides an upper bound on the actual payment, which ensures budget-feasibility, the bound may be loose and thus the mechanism creates budget surplus. In our mechanism, we adopt the bisection method to make full use of the budget.

III. Preliminaries

A. Statistical Interpolation - Kriging

Kriging [14] is a well-known geo-statistical interpolation technique, originally developed for mining and has been adapted to radio mapping [2]–[4]. For radio mapping, Kriging employs a Gaussian random field model for RSS at a point x

$$Z(x) = \mu(x) + \delta(x) \tag{1}$$

where $\mu(x)$ is the mean and $\delta(x)$ the residual at location x. The former captures path loss and shadowing at different locations, and the later represents possible sampling errors.

A fundamental function in this engine is the *semivariogram* $\gamma(\cdot)$, which models the variance between two points as a function of their distance. In practice, $\gamma(\cdot)$ is estimated from measurements and then fitted with parametric models such as spherical and exponential models. In *Ordinary Kriging* (OK), $Z(x)$ is assumed to be intrinsically stationary, that is,

$$\begin{aligned} \mathbb{E}[Z(x)] &= \mu(x) = \mu \\ \mathbb{E}[(Z(x_1) - Z(x_2))^2] &= 2\gamma(h) \end{aligned} \tag{2}$$

where μ is an unknown constant and $h = ||x_1 - x_2||$. The relationship between $\gamma(h)$ and the covariance function $C(h)$ is given by $C(h) = C(0) - \gamma(h)$. In this paper, we focus on OK due to its popularity[2].

Given a set of measurements $\mathcal{A}$ at locations $x_1, x_2, ..., x_n$, the predicted value at an unmeasured point x_0 is

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \omega_i \cdot Z(x_i) \tag{3}$$

where $\{\omega_i\}$ are normalized weights, i.e., $\sum_{i=1}^{n} \omega_i = 1$. Thus, Kriging produces a linear unbiased estimator. Minimization of the Mean Squared Error (MSE) $\mathbb{E}[(\hat{Z}(x_0) - Z(x_0))^2]$ with respect to $\{\omega_i\}$ under the normalization constraint leads straightforwardly to a set of linear equations (aka Kriging system). The optimal coefficients are given by $\omega^* = (\omega_i^*)_{i \in \mathcal{A}} = \Sigma_{\mathcal{A}\mathcal{A}}^{-1} \Sigma_{\mathcal{A}\mathbf{x}_0}$, where $\Sigma_{\mathcal{A}\mathcal{A}}$ is the covariance matrix, and $\Sigma_{\mathcal{A}\mathbf{x}_0}$ is the vector of cross-covariances between $\{Z(x_i)\}$ and $Z(x_0)$. The minimized MSE (also the *Kriging variance* (K-var) since the estimator is unbiased) is given by

$$\sigma^2_{\mathbf{x}_0|\mathcal{A}} = \sigma^2_{\mathbf{x}_0} - \Sigma^T_{\mathcal{A}\mathbf{x}_0} (\Sigma^{-1}_{\mathcal{A}\mathcal{A}})^T \Sigma_{\mathcal{A}\mathbf{x}_0} \tag{4}$$

[2]A generalized technique called *Universal Kriging* relieves the constant-mean assumption. In practice, it is common to first estimate and subtract $\mu(x)$ from $Z(x)$, which makes $\delta(x)$ an intrinsically stationary process with a constant (zero) mean and OK may be applied.

where $\sigma^2_{\mathbf{x}_0}$ is the K-var when $\mathcal{A} = \emptyset$[3]. K-var represents the prediction uncertainty at the unmeasured location and is often used as the estimator design metric (smaller K-Var implies better REM estimator).

B. Spatial Sampling Design

A classic spatial sampling problem discussed in [16] is the following. Consider a set of candidate sample locations $\mathcal{C}$ and a set of unmeasured locations $\mathcal{D}$, which are often a grid discretization of a continuous region. The task is to choose a subset $\mathcal{A} \subseteq \mathcal{C}$ up to k elements that minimizes the average K-var over $\mathcal{D}$, or equivalently, maximizes the average K-var reduction[4] $\phi(\mathcal{A})$, which is given by

$$\phi(\mathcal{A}) = \frac{1}{|\mathcal{D}|} \sum_{x_0 \in \mathcal{D}} \left(\sigma^2_{x_0} - \sigma^2_{x_0|\mathcal{A}} \right) \quad (5)$$

Since the problem $\max_{\mathcal{A} \subseteq \mathcal{C}, |\mathcal{A}| \le k} \phi(\mathcal{A})$ is NP-hard in general [18], computing the optimal solution is difficult. However, by exploiting structural properties of $\phi(\mathcal{A})$ such as *submodularity* and *monotonicity* [18], [19] with respect to the cardinality of $\mathcal{A}$, an *approximate* solution could be obtained with greedy algorithms.

Formally, a set function $f : 2^{\mathcal{C}} \to \mathcal{R}$ is called *submodular*, if $f(\mathcal{A} \cup \{x\}) - f(\mathcal{A}) \ge f(\mathcal{B} \cup \{x\}) - f(\mathcal{B})$ for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$ and any $x \in \mathcal{C} \setminus \mathcal{B}$. The notion of submodularity describes the *diminishing returns* behavior: adding a new element increases f more, if there are fewer elements so far, and less, if there are more elements. A set function $f : 2^{\mathcal{C}} \to \mathcal{R}$ is said to be *monotone*, if $f(\mathcal{A}) \le f(\mathcal{B})$ for any $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{C}$. In fact, $\phi(\mathcal{A})$ is also *non-negative*. It is well-known that for a submodular, monotone and non-negative function, a greedy algorithm finds an approximate solution guaranteed to be within $1 - 1/e$ of the optimal solution, and no polynomial-time algorithm can achieve a better guarantee unless P = NP [20].

C. Myerson's Characterization Theorem

In mechanism design, truthfulness means that it is a dominant strategy for users to report their true costs regardless of other users' bids [21]. Denote the cost, bid and payment of user i as c_i, b_i and p_i, respectively. A mechanism $\mathcal{M}$ consists of a *selection* rule $\chi : (\mathbf{b}, \mathbf{x}) \to 2^{[n]}$, and a *payment* rule $\psi : (\mathbf{b}, \mathbf{x}) \to \mathcal{R}^n_+$, where $\mathbf{b}$ and $\mathbf{x}$ are the bid and location vectors respectively. Since the cost is the only private parameter, the Myerson's characterization theorem [10], [22], which specifies the sufficient and necessary conditions for truthfulness, is very relevant.

Theorem 1 (Myerson's Characterization)**.** In single parameter domains, a mechanism $\mathcal{M} = (\chi, \psi)$ is truthful if and only if:

1) χ is *monotone*: $\forall i \in \Omega$, if $b'_i \le b_i$ then $i \in \chi(b_i, b_{-i}, \mathbf{x})$ implies $i \in \chi(b'_i, b_{-i}, \mathbf{x})$ for any given b_{-i}. That is, a winner keeps winning if it unilaterally decreases his bid.
2) ψ pays winners the *threshold amounts*: $p_i = \sup\{b_i : i \in \chi(b_i, b_{-i}, \mathbf{x})\}$ for any given b_{-i} [5]. That is, the payment is the maximum bid that still wins.

A simple example is the single-good second-price (reverse) auction with private costs [21], where the user with the lowest bid is the only winner (selection rule), who is paid the second lowest bid (payment rule). We may verify its truthfulness with Myerson's characterization. First, the selection rule is monotone, that is, the winner still wins if it submits an even lower bid. Second, the winner is paid the threshold amount: if it submits a bid higher than the second lowest bid, it is no longer the winner. Hence, the second-price auction is truthful.

IV. System Model

Fig. 1 illustrates our crowdsourcing system for radio mapping, which consists of a centralized server called *platform*, and CR equipped *users*, who are spatially distributed and connected to the platform. We assume that each user knows its current geo-location with high accuracy and is able to collect high-quality[6] spectrum data.

The platform seeks to acquire sensing data from users periodically for the spectrum band of interest. At the beginning of a period, the platform announces a sensing request *without* specific location tags, which contains detailed sensing instructions (such as center frequency, sampling rate etc.). Each user i in the desired region D can compete for the task, and incurs a privately known *cost* $c_i > 0$ for sensing. We assume *no entry or other overhead costs*, that is, a user does not incur a fee to bid nor does it pay to communicate with the platform. User i submits its location x_i and a *bid* $b_i \ge c_i$, the minimum payment it is willing to accept. Denote the set of bidders as $\Omega = \{1, 2, ..., n\}$ at locations $\{x_1, x_2, ..., x_n\}$, where $n \ge 2$ and $x_i \in D$. Upon receiving a bid-location profile $(\mathbf{b}, \mathbf{x})$ where $\mathbf{b} = (b_1, b_2, ..., b_n)$ and $\mathbf{x} = (x_1, x_2, ..., x_n)$, the platform selects a winner subset $\mathcal{A} \subseteq \Omega$ and determines the payment $p_i > 0$ for each winner $i \in \mathcal{A}$ ($p_i = 0$ for $i \notin \mathcal{A}$). Finally, it collects sensing data from winners.

We assume that users are rational and make decisions in their best interest. Each user i has a *utility* of $p_i - c_i$ if selected, and 0 otherwise. We are interested in the *strategic* case, where each user wants to maximize its own utility by taking strategic actions, for example, submitting a bid (possibly) much higher than its true cost. In addition, we assume that users are non-collaborative and honest in following the protocol.

[3]Strictly speaking, $\sigma^2_{\mathbf{x}_0}$ is undefined according to Eq. 4 when $|\mathcal{A}| = 0$; when $|\mathcal{A}| = 1$, it depends on both $C(0)$ and μ, which is unknown. Due to this reason, *GeoR* [15], the widely-used geo-statistics library in R, requires at least two data points for Kriging. In our implementation, when $\mathcal{A} = \emptyset$ (no data points in the interested region), we bypass the issue by introducing two data points that are very far away from the target region; for $|\mathcal{A}| = 1$, we simply add a second data point very close to the existing one.

[4]An alternative metric is the *mutual information* [17]. Since the two criteria share very similar properties such as submodularity and monotonicity, we focus on the average K-var reduction in this work. However, our discussion also extends to the mutual information criterion.

[5]$b_{-i} = (b_j)_{j \ne i,\ j \in \Omega}$ are other users' bids.

[6]Data quality depends on various factors in practice such as noise power and local environment (e.g., indoor/outdoor). In this paper, we assume that CR devices (or antennas) are located outdoors, whose noise power are low enough to reliably sense the presence of primary signals in bands of interest. For instance, if a user wants to compete for sensing tasks in TV bands, its CR device should be able to sense as low as the service threshold of -84 dBm over a 6-MHz channel for full-power digital TV [1]. To enforce this assumption, the platform may ask users to report their noise levels and local environments, and only allow qualified users to compete for a particular task.

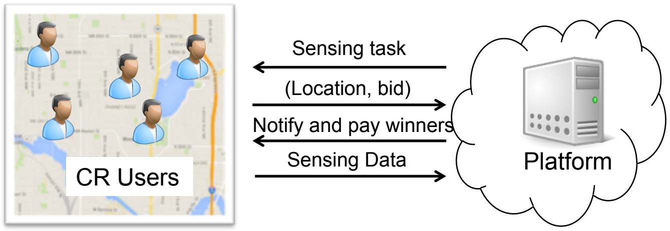


Fig. 1: Incentivized crowdsourcing system for radio mapping.

Considerations of security and privacy enhancement within this framework is left as future work.

The platform aims to maximize $\phi(\mathcal{A})$[7] for a given budget B. We assume that $\gamma(h)$ or $C(h)$ is already estimated a-priori and thus known to the platform for the current period. The platform's task is to design a *mechanism* $\mathcal{M} = (\chi, \psi)$ with the following desirable properties:

- **Computational efficiency**: selection and payment rules can be computed in polynomial time[8].
- **Individual rationality**: each participating user will have a non-negative utility, i.e., $p_i \geq b_i$ for $i \in \mathcal{A}$.
- **Budget feasibility**: the total amount of payments does not exceed a given budget, i.e., $\sum_{i \in \mathcal{A}} p_i \leq B$.
- **Truthfulness**: it is a dominant strategy for users to report their true costs regardless of other users' bids.

The first three properties characterize a feasible crowdsourcing system in a real world environment while the final property counteracts the possibility of market manipulation and strategizing.

V. Incentive Mechanism Design

Similar to [10], [11], Theorem 1 is exploited to design a truthful mechanism. We first consider an auction-based budget-free mechanism with a cardinality constraint, and then propose a budget-feasible version that enforces the budget constraint by translating it to an appropriately modified cardinality constraint. Finally, we show that the proposed mechanism is computationally efficient, individually rational, budget feasible and truthful, and illustrate it with an example.

A. Budget-Free Mechanism with a Cardinality Constraint

In the budget-free case, the platform takes a bid-location profile $(\mathbf{b}, \mathbf{x})$ and a cardinality constraint k, and tries to solve the following problem:

$$\max_{\mathcal{A} \subseteq \Omega} \ \phi(\mathcal{A})$$
$$\text{subject to } |\mathcal{A}| \leq k \tag{6}$$

Since both individual rationality and truthfulness will be enforced in payment determination using Theorem 1, they are not imposed as explicit constraints. It would seem that the above is the classical spatial sampling problem, and a greedy algorithm that only considers user locations and iteratively selects users with the maximum marginal contribution is the solution. However, since the above selection rule does not depend on users' bids, a winner can submit an arbitrarily high bid and still win. Thus, to ensure truthfulness, we need to consider *location and bid jointly* for winner selection. Similar to [10], [11], we use the *normalized marginal contribution* (i.e., marginal contribution divided by their bids) as the metric, which yields a monotone selection rule, as we will show later.

Algorithm 1 describes the budget-free mechanism. As we can see, the selection rule (Lines 4-7) is based on the greedy heuristic that selects winners one by one iteratively according to their normalized marginal contribution until k winners are selected. Consider n users in Ω, labeled from 1 to n. In the j-th iteration, the current set of $(j-1)$ winners is denoted by $\mathcal{A}_{j-1}$, where $j \geq 1$ and $\mathcal{A}_0 = \emptyset$. The marginal contribution of each user $i \in \Omega \setminus \mathcal{A}_{j-1}$ is given by: $m_{\mathcal{A}_{j-1}}(i) = \phi(\mathcal{A}_{j-1} \cup \{i\}) - \phi(\mathcal{A}_{j-1})$. Define $[j] = \arg\max_{i \in \Omega \setminus \mathcal{A}_{j-1}} \frac{m_{\mathcal{A}_{j-1}}(i)}{b_i}$, which is the index of the j-th winner over Ω. To simplify notation, we write $m_{[j]}$ instead of $m_{\mathcal{A}_{j-1}}([j])$. Note that $\phi(\mathcal{A}_j) = \sum_{i \leq j} m_{[i]}$ for all $j \leq k$. The submodularity of $\phi(\mathcal{A})$ implies that

$$m_{[1]} \geq m_{[2]} \geq ... \geq m_{[k]} \tag{7}$$

and the selection order implies that

$$\frac{m_{[1]}}{b_{[1]}} \geq \frac{m_{[2]}}{b_{[2]}} \geq ... \geq \frac{m_{[k]}}{b_{[k]}} \tag{8}$$

We now show that Eq. (8) is true by contradiction. Consider winners $[i]$ and $[j]$, where $i < j$. Suppose that $\frac{m_{[i]}}{b_{[i]}} < \frac{m_{[j]}}{b_{[j]}}$. Denote the marginal contribution of the j-th winner in the i-th iteration as $m'_{[j]}$. In the presence of submodularity, we have $m'_{[j]} = m_{\mathcal{A}_{i-1}}([j]) \geq m_{[j]} = m_{\mathcal{A}_{j-1}}([j])$, since $\mathcal{A}_{i-1} \subseteq \mathcal{A}_{j-1}$. Thus, it holds that $\frac{m_{[i]}}{b_{[i]}} < \frac{m_{[j]}}{b_{[j]}} \leq \frac{m'_{[j]}}{b_{[j]}}$. In other words, the j-th winner would have been selected earlier in the i-th iteration, which is a contradiction with our assumption. Therefore, Eq. (8) is true.

Next comes the payment determination. The key is to find the maximum bid each winner can submit that allows it to win. The corresponding pseudo-codes are in Lines 10-18. Consider the i-th winner among Ω, denoted as $[i]$. Define a new set $\Omega' = \Omega \setminus \{[i]\}$. Similar to Eq. (8), sort users in Ω' according to their normalized marginal contribution. Denote the first $(j-1)$ winners among Ω' as $\mathcal{A}'_{j-1}$, and the index of the j-th winner as $[j]'$. In order for winner $[i]$ to replace winner $[j]'$, its normalized marginal contribution needs to be larger than that of winner $[j]'$, i.e.,

$$\frac{m_{\mathcal{A}'_{j-1}}([j]')}{b_{[j]'}} < \frac{m_{\mathcal{A}'_{j-1}}([i])}{b_{[i]}} \tag{9}$$

[7] The criterion of "average" K-var reduction implicitly assumes equally important subregions with the same accuracy requirement. In practice, accuracy requirement may vary over subregions, and those that require greater interpolation accuracy will need more samples in general. To achieve this goal, our proposed framework can be readily extended in the following two ways. First, the platform may assign larger budgets to more important subregions and smaller budgets to others, given the same budget in total.

Alternatively, the platform may adopt a "weighted" criterion and assign more weights to unmeasured locations from subregions that require greater accuracy. By doing so, the platform tends to select more samples from those subregions in the processing of maximizing the "weighted" phi(A).

[8] The computational complexity is evaluated in terms of the number of calls to $\phi(\mathcal{A})$, which can be computed in polynomial time.

So the maximum bid (or the conditional threshold payment) for winner $[i]$ that allows it to replace winner $[j]'$ is

$$p_{[i],[j]'} = \frac{m_{\mathcal{A}'_{j-1}}([i])}{m_{\mathcal{A}'_{j-1}}([j]')} \cdot b_{[j]'} \tag{10}$$

Note that as long as winner $[i]$ is ahead of any winner $[j]' \in \Omega'$ in terms of ordering, where $j \leq k$, it is guaranteed that the i-th winner still wins. Therefore, the maximum bid (i.e., the threshold payment) for the i-th winner is

$$p_{[i]} = \max_{1 \leq j \leq k} \{p_{[i],[j]'}\}, \; i = 1, 2, ..., k \tag{11}$$

Algorithm 1: `budget_free_mechanism(`$\mathbf{b}, \mathbf{x}, k$`)`

input : $\mathbf{b}, \mathbf{x}, k$
output: $\mathcal{A}$ – selected subset, $\mathbf{p}$ – payment vector

$\Omega \leftarrow \{1, 2, ..., n\}$ // Or $\Omega \leftarrow \{x_1, x_2, ..., x_n\}$
// Selection Rule
$\mathcal{A} \leftarrow \emptyset, \mathcal{U} \leftarrow \Omega$;
while $\mathcal{U} \neq \emptyset$ **and** $|\mathcal{A}| < k$ **do**
    $j \leftarrow \arg\max_{i \in \mathcal{U}} \left(\phi(\mathcal{A} \cup \{i\}) - \phi(\mathcal{A})\right) / b_i$;
    $\mathcal{A} \leftarrow \mathcal{A} \cup \{j\}, \mathcal{U} \leftarrow \mathcal{U} \setminus \{j\}$
// Payment Rule
foreach $i \in \Omega$ **do** $p_i \leftarrow 0$;
foreach $i \in \mathcal{A}$ **do**
    $\Omega' \leftarrow \Omega \setminus \{i\}$;
    $\mathcal{A}' \leftarrow \emptyset, \mathcal{U} \leftarrow \Omega'$;
    while $\mathcal{U} \neq \emptyset$ **and** $|\mathcal{A}'| < k$ **do**
        $j \leftarrow \arg\max_{l \in \mathcal{U}} \left(\phi(\mathcal{A}' \cup \{l\}) - \phi(\mathcal{A}')\right) / b_l$;
        $m'_j \leftarrow \phi(\mathcal{A}' \cup \{j\}) - \phi(\mathcal{A}')$;
        $\tilde{m}_i \leftarrow \phi(\mathcal{A}' \cup \{i\}) - \phi(\mathcal{A}')$;
        $p_i \leftarrow \max\{p_i, \frac{\tilde{m}_i}{m'_j} \cdot b_j\}$;
        $\mathcal{A}' \leftarrow \mathcal{A}' \cup \{j\}, \mathcal{U} \leftarrow \mathcal{U} \setminus \{j\}$;
return $(\mathcal{A}, \mathbf{p})$;

B. Budget-Feasible Mechanism

We now consider the budget constraint. Denote the problem in Eq. (6) and the corresponding winner set as $H(\mathbf{b}, \mathbf{x}, k)$ and $\mathcal{A}_k$ respectively. The total payment is $P_{total}(k) = \sum_{i \in \mathcal{A}_k} p_i$, which is a function of k. Then the budget-feasible version aims to solve the following problem:

$$\begin{aligned} &\max_{k} \; H(\mathbf{b}, \mathbf{x}, k) \\ &\text{subject to } P_{total}(k) \leq B \end{aligned} \tag{12}$$

Suppose that $k_1 < k_2$ and denote the corresponding winner sets as $\mathcal{A}^{(1)}$ and $\mathcal{A}^{(2)}$. It is easy to see that $\mathcal{A}^{(1)} \subset \mathcal{A}^{(2)}$ because both adopt the same selection rule. Since $\phi(\mathcal{A})$ is a monotonically increasing function, larger k results in larger $\phi(\mathcal{A})$. Thus, Eq. (12) is equivalent to maximizing k such that $P_{total}(k) \leq B$. One simple way is to enumerate every k value in order until the total payment exceeds the budget. Alternatively, the *bisection* method may be used to speed up the search process by leveraging the fact that $P_{total}(k)$ is a monotonically increasing function of k.

Lemma 1. *The total payment of the budget-free mechanism is a monotonically increasing function of k.*

Proof: Suppose $k_1 < k_2$. Let the outputs of the budget-free mechanism given k_1 and k_2 be $(\mathcal{A}^{(1)}, \mathbf{p}^{(1)})$ and $(\mathcal{A}^{(2)}, \mathbf{p}^{(2)})$ respectively. As in Eq. (8), we sort users according to their normalized marginal contribution, and denote the i-th winner as $[i]$. Then, we have $\mathcal{A}^{(1)} = \{[1], [2], ..., [k_1]\} \subset \mathcal{A}^{(2)} = \{[1], [2], ..., [k_2]\}$. Denote the payments as $\{p^{(1)}_{[i]} : i = 1, 2, ..., k_1\}$ and $\{p^{(2)}_{[i]} : i = 1, 2, ..., k_2\}$. For any $i \leq k_1$, it holds that

$$\begin{aligned} p^{(2)}_{[i]} &= \max\{\max_{1 \leq j \leq k_1} p_{[i],[j]'}, \max_{k_1 < j \leq k_2} p_{[i],[j]'}\} \\ &= \max\{p^{(1)}_{[i]}, \max_{k_1 < j \leq k_2} p_{[i],[j]'}\} \geq p^{(1)}_{[i]} \end{aligned} \tag{13}$$

For any $i \in (k_1, k_2]$, $p^{(2)}_{[i]} > 0$. Thus, $\sum_{1 \leq i \leq k_2} p^{(2)}_{[i]} > \sum_{1 \leq i \leq k_1} p^{(1)}_{[i]}$. ■

The budget-feasible mechanism is described in Algorithm 2. The bisection method requires a lower bound l for k such that the budget constraint can be met, and an upper bound u such that the budget constraint cannot be met. A tolerance value ϵ is also needed as the stopping condition. We assume that the budget is neither too small nor too large, so that the winning set contains at least one sample but not all samples are affordable. Thus, let $l = 1$, $u = n$ and $\epsilon = 1$ as in Line 1. Then, in the while loop (Lines 2-10), the midpoint between u and l is fed to the budget-free mechanism to check the budget constraint: if yes, the midpoint becomes the new lower bound, and the new upper bound otherwise. The number of calls to the budget-free mechanism is $\log_2((u - l)/\epsilon) \approx \log_2(n)$.

Algorithm 2: `budget_feasible_mechanism(`$\mathbf{b}, \mathbf{x}, B$`)`

input : $\mathbf{b}, \mathbf{x}, B$
output: $\mathcal{A}$ – selected subset, $\mathbf{p}$ – payment vector

1 $l \leftarrow 1$, $u \leftarrow n$, $\epsilon \leftarrow 1$;
2 **while** $u - l > \epsilon$ **do**
3     $k \leftarrow \lfloor (u - l)/2 \rfloor$;
4     $(\mathcal{A}', \mathbf{p}') \leftarrow$ `budget_free_mechanism(`$\mathbf{b}, \mathbf{x}, k$`)`;
5     $P_{total} \leftarrow \sum_{i \in \mathcal{A}'} p'_i$;
6   **if** $P_{total} \leq B$ **then**
7         $l \leftarrow k$;
8         $\mathcal{A} \leftarrow \mathcal{A}'$, $\mathbf{p} \leftarrow \mathbf{p}'$;
9   **else**
10         $u \leftarrow k$
11 **return** $(\mathcal{A}, \mathbf{p})$;

C. Analysis

Now we prove the computational efficiency, individual rationality and truthfulness of the proposed budget-feasible mechanism.

Lemma 2. *The proposed budget-feasible mechanism is computationally efficient (i.e., in polynomial time).*

Proof: For the budget-free mechanism, the complexity of the greedy selection is $\mathcal{O}(k \cdot n)$, since finding the user with maximum normalized marginal contribution takes $\mathcal{O}(n)$ time and we need to find k such users. In payment determination, the greedy process is executed repeatedly to determine each winner's payment. So the complexity is $\mathcal{O}(k^2 \cdot n)$ for the budget-free mechanism. Since k is bounded by n, the overall complexity for a general k is bounded by $\mathcal{O}(n^3)$.

Since the budget-feasible mechanism adopts the bisection method which requires $\log_2(n)$ calls to the budget-free mechanism in the worst case, the complexity then becomes $\mathcal{O}(n^3 \log_2 n)$, which is polynomial-time. ■

Lemma 3. *The proposed budget-feasible mechanism is individually rational.*

Proof: Since $p_{[i]} \geq p_{[i],[j]'}$ for $i = 1, 2, ..., k$, it suffices to prove that $p_{[i],[j]'} \geq b_{[i]}$ for some j. Observe that the first $(i-1)$ winners over Ω are the same with those over Ω', i.e., $\mathcal{A}'_{i-1} = \mathcal{A}_{i-1}$. Due to the absence of winner $[i]$ in Ω', some other user now becomes winner $[i]'$, who ranks behind winner $[i]$ in the original set when it is present, that is,

$$\frac{m_{\mathcal{A}_{i-1}}([i]')}{b_{[i]'}} \leq \frac{m_{\mathcal{A}_{i-1}}([i])}{b_{[i]}}$$

$$\Rightarrow \frac{m_{\mathcal{A}'_{i-1}}([i]')}{b_{[i]'}} \leq \frac{m_{\mathcal{A}'_{i-1}}([i])}{b_{[i]}} \qquad \text{(Since } \mathcal{A}'_{i-1} = \mathcal{A}_{i-1})$$

$$\Rightarrow p_{[i],[i]'} = \frac{m_{\mathcal{A}'_{i-1}}([i])}{m_{\mathcal{A}'_{i-1}}([i]')} \cdot b_{[i]'} \geq b_{[i]} \qquad (14)$$

Hence, it holds that $p_{[i]} \geq p_{[i],[i]'} \geq b_{[i]}$ for $i = 1, 2, ..., k$. ■

Lemma 4. *The proposed budget-feasible mechanism is truthful.*

Proof: First, we prove that the selection rule is monotone. Consider the i-th winner, denoted as $[i]$. If winner $[i]$ announces $b'_{[i]} < b_{[i]}$, it holds that $\frac{m_{[i]}}{b'_{[i]}} > \frac{m_{[i]}}{b_{[i]}} \geq \frac{m_{[k]}}{b_{[k]}}$ and winner $[i]$ still wins. In other words, bidding a smaller value cannot push winner $[i]$ backwards in the sorting. Hence, the selection rule is monotone.

Then, we prove that the payments are threshold amounts. Assume that winner $[i]$ announces a bid $b_{[i]} > p_{[i]}$. By definition, we know that $b_{[i]} > p_{[i],[j]'}$ for all $1 \leq j \leq k$ and $b_{[i]} > p_{[i],[k]'}$ in particular. It means that when winner $[i]$ bids this amount, it will not be placed ahead of the k-th winner, even if it is included in the sorting again. Thus, winner $[i]$ cannot win by bidding $b_{[i]} > p_{[i]}$, and $p_{[i]}$ the threshold payment to winner $[i]$.

By invoking Theorem 1, it holds that the proposed budget-feasible mechanism is truthful. ■

The above lemmas prove the following theorem.

Theorem 2. *The proposed budget-feasible mechanism is computationally efficient, individually rational and truthful.*

D. An Example

For illustration of the proposed algorithm, we consider a simple example shown in Fig. 2. The square presents the region of interest for radio mapping, which is discretized into a grid of 9 locations of interest. There are four interested users located at x_1, x_2, x_3, x_4, who are labeled as $1, 2, 3$ and 4. Suppose that they bid $0.1, 0.2, 0.3, 0.4$, respectively and the budget is 0.5. The semivariogram $\gamma(h)$ is a spherical model, given by $\gamma(h) = a + (s-a)\left(\frac{3}{2}(\frac{h}{r}) - \frac{1}{2}(\frac{h}{r})^3\right)$ for $0 \leq h \leq r$, and $\gamma(h) = s$ for $h > r$, where $a = 0, s = 5$ and $r = 3$. Values of $\phi(\mathcal{A})$ are given in Table I.

$\mathcal{A}$	$\phi(\mathcal{A})$	$\mathcal{A}$	$\phi(\mathcal{A})$
$\emptyset$	0	$\{2,3\}$	6.38
$\{1\}$	4.34	$\{2,4\}$	5.99
$\{2\}$	4.29	$\{3,4\}$	5.23
$\{3\}$	4.29	$\{1,2,3\}$	7.03
$\{4\}$	4.55	$\{1,2,4\}$	6.89
$\{1,2\}$	6.00	$\{1,3,4\}$	6.54
$\{1,3\}$	6.04	$\{2,3,4\}$	6.55
$\{1,4\}$	6.22	$\{1,2,3,4\}$	7.20

TABLE I: Evaluated $\phi(\mathcal{A})$ for any $\mathcal{A} \subseteq \{1,2,3,4\}$.

We first demonstrate how the budget-free mechanism works. Observe that the greedy selection order by the normalized marginal contribution is given by: $1 \to 2 \to 3 \to 4$. The process of computing threshold payments for a single and two winners (i.e., $k = 1$ and 2) is given as follows.

For $k = 1$, we have $[1] = 1$ (the only winner is 1):

- In the absence of $[1]$, the only winner is $[1]' = 2$
  $\mathcal{A}'_0 = \emptyset$, $p_{[1],[1]'} = \frac{m_{\mathcal{A}'_0}([1])}{m_{\mathcal{A}'_0}([1]')} \cdot b_{[1]'} = \frac{4.34}{4.29} \cdot 0.2 \approx 0.202$
  $p_{[1]} = p_{[1],[1]'} = 0.202$;
- Hence, the payment to the only winner is 0.202.

For $k = 2$, we have $[1] = 1$ and $[2] = 2$:

- In the absence of $[1]$, winners are $[1]' = 2$ and $[2]' = 3$.
  1) $\mathcal{A}'_0 = \emptyset$, $p_{[1],[1]'} = \frac{m_{\mathcal{A}'_0}([1])}{m_{\mathcal{A}'_0}([1]')} \cdot b_{[1]'} = \frac{m_\emptyset(1)}{m_\emptyset(2)} \cdot b_2 = \frac{4.34}{4.29} \cdot 0.2 \approx 0.202$;
  2) $\mathcal{A}'_1 = \{[1]'\} = \{2\}$, $p_{[1],[2]'} = \frac{m_{\mathcal{A}'_1}([1])}{m_{\mathcal{A}'_1}([2]')} \cdot b_{[2]'} = \frac{m_{\{2\}}(1)}{m_{\{2\}}(3)} \cdot b_3 = \frac{6.00-4.29}{6.38-4.29} \cdot 0.3 \approx 0.245$;
  3) $p_{[1]} = \max\{p_{[1],[1]'}, p_{[1],[2]'}\} = 0.245$;
- In the absence of $[2]$, winners are $[1]' = 1$ and $[2]' = 3$.
  1) $\mathcal{A}'_0 = \emptyset$, $p_{[2],[1]'} = \frac{m_{\mathcal{A}'_0}([2])}{m_{\mathcal{A}'_0}([1]')} \cdot b_{[1]'} = \frac{m_\emptyset(2)}{m_\emptyset(1)} \cdot b_1 = \frac{4.29}{4.34} \cdot 0.1 \approx 0.099$;
  2) $\mathcal{A}'_1 = \{[1]'\} = \{1\}$, $p_{[2],[2]'} = \frac{m_{\mathcal{A}'_1}([2])}{m_{\mathcal{A}'_1}([2]')} \cdot b_{[2]'} = \frac{m_{\{1\}}(2)}{m_{\{1\}}(3)} \cdot b_3 = \frac{6.00-4.34}{6.04-4.34} \cdot 0.3 \approx 0.293$;
  3) $p_{[2]} = \max\{p_{[2],[1]'}, p_{[2],[2]'}\} = 0.293$;
- Hence, the payments to winners $[1] = 1$ and $[2] = 2$ are 0.245 and 0.293 respectively, and the total is 0.538.

Now we consider the budget-feasible mechanism. The bisection method is first initialized with $l = 1$ and $u = 4$. Then it checks the budget feasibility when $k = \lfloor (4+1)/2 \rfloor = 2$, and realizes that the budget constraint cannot be met (as shown above). So it sets $u = 2$. Now since the gap between the lower and upper bounds is within the tolerance (i.e., $\epsilon = 1$), the mechanism returns the result when $k = l$, which is a budget-feasible solution.

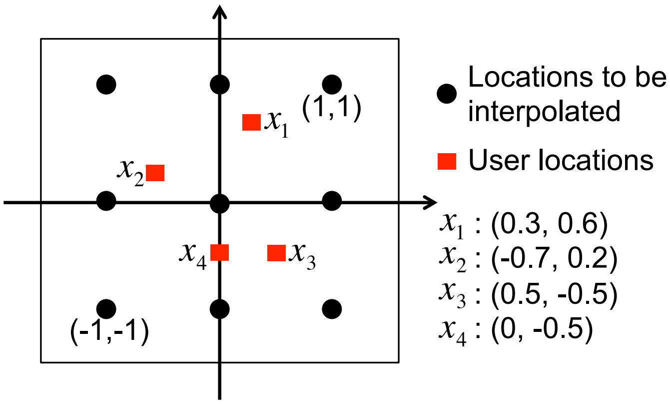


Fig. 2: Topology of the example with user locations and locations to be interpolated. The K-var reduction is averaged over the 9 locations of interest to obtain $\phi(\mathcal{A})$.

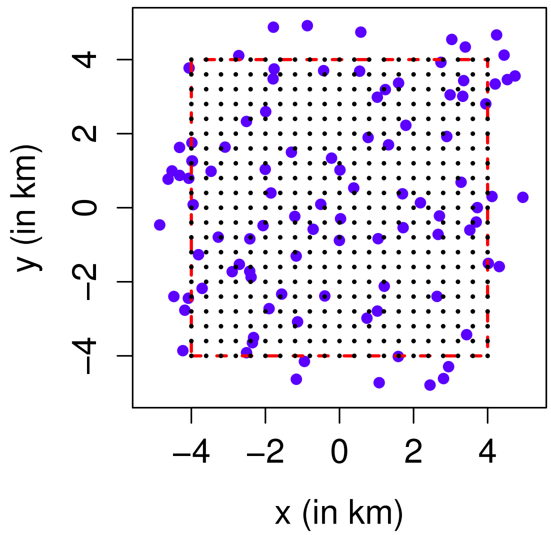


Fig. 3: Sample topology. There are 100 randomly distributed users (in blue dots) over the 10km-by-10km region. The region of interest is the inner red square, which is discretized into a total of 121 locations (in black dots).

VI. Performance Evaluation

In this section, we evaluate the performance of proposed mechanisms and compare them against the baseline mechanism in [11]. For convenience, we use the abbreviations *BFreeMech* and *BFeaMech* for the budget-free and budget-feasible mechanisms respectively.

A. Baseline Mechanism

In [11], authors proposed a randomized budget-feasible mechanism for general submodular monotone functions, which is computationally efficient, individually rational and truthful.

There are two main differences between their mechanism and ours. First, their mechanism is randomized: with a certain probability (i.e., 0.4), it returns a single user with the maximum marginal contribution (unnormalized) and pays it B; otherwise, it runs a greedy scheme which select multiple users and determines payments based on Myerson's characterization. The logic behind the randomness is the following. In some extreme cases, there exists some user with very large marginal contribution and very high cost. As a result, it will never be selected by a greedy algorithm, which yields unbounded performance. Hence, authors adopted the randomized approach in order to derive a certain performance bound.

Second, in their greedy scheme, the budget constraint is enforced through the proportional share allocation condition. Specifically, the greedy scheme selects the user with largest normalized marginal contribution in the j-th iteration only if $b_{[j]} \leq \frac{B}{2} \frac{m_{[j]}}{\sum_{i \in \mathcal{A}_{j-1} \cup \{[j]\}} m_i}$, and stops otherwise. This condition ensures that the final payment to winner $[i]$ in the winner set $\mathcal{A}_k$ is bounded above by $\frac{m_{[i]}}{\phi(\mathcal{A}_k)} \cdot B$; then the total payment will be bounded by B since $\frac{\sum_{i \leq k} m_{[i]}}{\phi(\mathcal{A}_k)} \cdot B = B$.

Since it is very unlikely that two or more samples would lead to lower $\phi(\mathcal{A})$ than a single sample in radio mapping, we only consider their greedy scheme as the baseline mechanism.

B. Simulation Setup

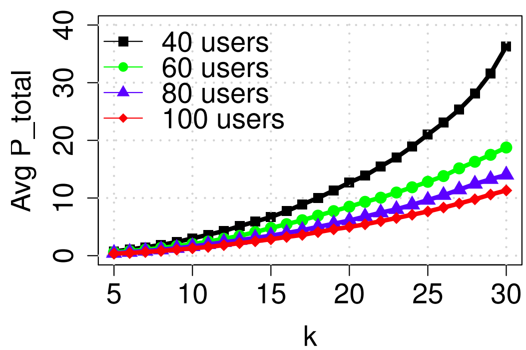
A sample topology with 100 users is given in Fig. 3, whose locations are randomly generated from the spatial Poisson random process within a 10km-by-10km region. The actual region of interest is the 8km-by-8km square, which is discretized into a grid of 121 locations of interest. This is because in practice estimation errors tend to be dominated by the border areas [23] and hence we focus on radio mapping for the inner region. User costs are i.i.d. random variables drawn from the uniform distribution over $[0, \kappa_{max}]$. Since the normalized marginal contribution is used as the metric in our proposed mechanism, the scale of κ_{max} has no impact on winner selection and payments are proportional to κ_{max}. Without loss of generality, we set $\kappa_{max} = 1$. The semivariogram $\gamma(h)$ is an exponential model obtained from a real-world measurement campaign in a suburban area [3], which is given by $\gamma(h) = a + (s-a)\left(1 - e^{-3h/r}\right)$, where $a = 6.48$, $s = 22.02$ and $r = 2.11$. Mechanisms are implemented in R [24] using the *geoR* [15] package.

C. Evaluation of BFreeMech

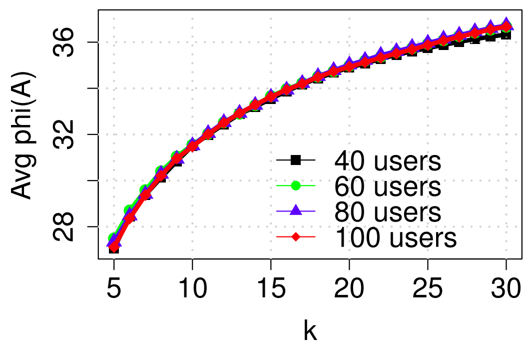
As discussed in Section V, BFeaMech's budget constraint is closely related to BFreeMech's cardinality constraint. Hence, we are interested in understanding the impact of the cardinality constraint on P_{total} and $\phi(\mathcal{A})$ for BFreeMech. In addition, we quantify the payment overhead used to ensure truthfulness, which is the difference between the total payment and the total amount of bids.

In this simulation, we first generated 30 sets of random costs and locations for 100 users. In each experiment, a particular number of users were randomly sampled from the 100 users; for the same set of users, BFreeMech was executed with a cardinality constraint varied from 5 to 30. Results were averaged over 30 experiments and are shown in Fig. 4.

1) Impact of k on P_{total} : There are two observations from Fig. 4a. First, the average P_{total} is monotonically increasing, and tends to increase at a faster rate as k increases. This could



(a)



(b)

Fig. 4: Impact of the cardinality constraint k on P_{total} and $\phi(\mathcal{A})$. (a) Average P_{total} as a function of k for different numbers of users. (b) Average $\phi(\mathcal{A})$ as a function of k for different numbers of users.

be explained as follows: in order to select one more winner, the payment for each existing winner is very likely to increase, which leads to a larger increase in P_{total} on average.

Second, for a relatively small k (e.g., less than 15) compared to the number of users n, we do not observe a significant difference in the average total payment. When k becomes larger, the average P_{total} tends to be inversely proportional to the number of participating users. For instance, when $k = 25$, the average total payment is 21.0 when there are 40 users, which is almost halved (i.e., 9.6) with a double number of users (i.e., 80 users).

2) Impact of k on $\phi(\mathcal{A})$: As shown in Fig. 4b, the average $\phi(\mathcal{A})$ increases as k increases, but the increasing rate keeps decreasing, which is mainly due to the submodularity property of $\phi(\mathcal{A})$. On the other hand, unlike P_{total}, the number of users has little effect on the average value of $\phi(\mathcal{A})$. It implies that to achieve certain performance, a corresponding number of samples need to be purchased, and with more participating users, the total payment will be lower.

3) Impact of k on payment overhead: As mentioned earlier, the actual payment made to a winner is always higher than its bid to ensure individual rationality, and more importantly truthfulness. Thus, for a set $\mathcal{A}$ of k winners, we may define the *payment overhead ratio* α as below

$$\alpha = \frac{\sum_{i\in\mathcal{A}} p_i - \sum_{i\in\mathcal{A}} b_i}{\sum_{i\in\mathcal{A}} b_i} \tag{15}$$

As illustrated in Fig. 5, when k is relatively small compared

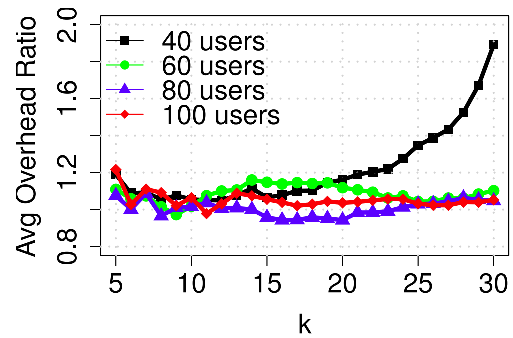


Fig. 5: Average payment overhead ratio as a function of k for different numbers of users.

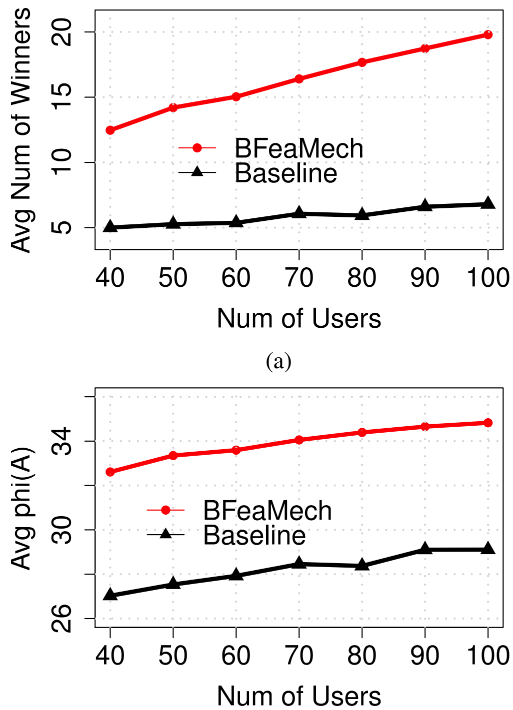
to n, α varies between 0.9 and 1.2 for different values of n. In other words, the platform needs to pay roughly double the total amount of bids to ensure truthfulness. On the other hands, when k is closer to n (e.g., more than 20 winners out of 40 users), α tends to increase rapidly. It implies that the platform may need to limit the number of winners depending on the number of participants, so as to avoid payment overhead.

D. Comparison with Baseline Mechanism

Now we compare the performance of BFeaMech against the baseline mechanism. For fair comparison, the same set of user costs and locations was fed to both mechanisms, and results were averaged over 30 experiments.

1) Impact of Number of Users: Fig. 6 illustrates the impact of n. We set the budget constraint to be 5. As we can see in Fig. 6a, the average number of purchased samples tends to increase linearly as a function of n for both mechanisms, but the slope of BFeaMech is greater than that of the baseline mechanism. As a result, BFeaMech performs much better than the baseline mechanism in terms of average $\phi(\mathcal{A})$ for the same n, with an improvement of 19.1%-21.2% for different numbers of users, as shown in Fig. 6b.

2) Impact of Budget: Fig. 7 illustrates the impact of B on the average number of samples and $\phi(\mathcal{A})$ for both mechanisms. We set the number of users to be 100. As shown in Fig. 7a, we can observe that the average number of purchased samples increases as B increases for both mechanisms, but it grows much faster for BFreaMech compared to the baseline mechanism. The reason is as follows. Suppose that k winners are selected by the baseline mechanism. It holds that $b_{[k+1]} > \frac{B}{2}\frac{m_{[k+1]}}{\sum_{i\in\mathcal{A}_k\cup\{[k+1]\}} m_i}$, or equivalently, $\frac{m_{[k+1]}}{b_{[k+1]}} < 2\frac{\sum_{i\in\mathcal{A}_k\cup\{[k+1]\}} m_i}{B}$. In order to get the $(k+1)$-th winner, the budget has to increase in a way so that the right-hand side is less or equal to the left-hand side. If we plot the average budget as a function of the number of winners k for the baseline mechanism (Fig. 8), we can see that it tends to increase exponentially as k increases, and thus the additional budget needed to get one more winner keeps increasing. Also notice that the total payment to k winners made by BFreeMech is much less than the budget required the baseline mechanism. Since both mechanisms determine payments in a very similar manner, it implies that the baseline mechanism does not make



(a)

(b)

Fig. 6: Impact of number of users n on the performance of BFeaMech and the baseline mechanism. (a) Average number of winners or purchased samples as a function of n. (b) Average $\phi(\mathcal{A})$ as a function of n. BFeaMech is able to consistently achieve an average value of $\phi(\mathcal{A})$ 19.1%-21.2% higher than that of the baseline mechanism for different numbers of users.

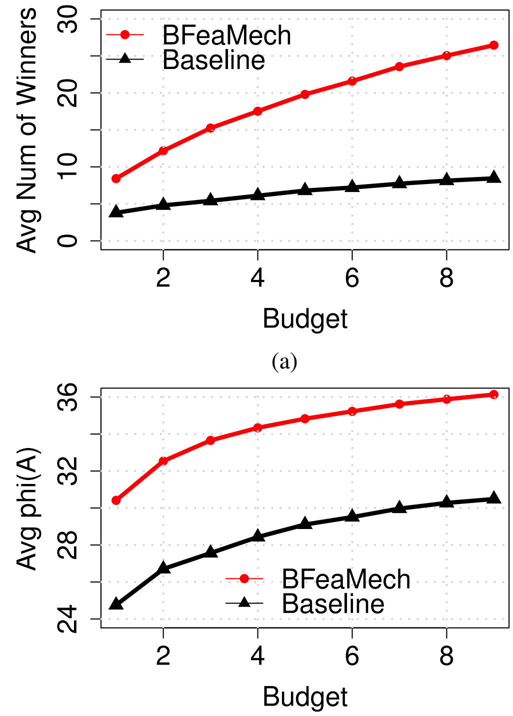
full use of the budget, and with a larger the budget, the budget surplus will also become greater. Therefore, BFreaMech tends to make better use of the additional budget and is able to purchase more samples than the baseline mechanism in radio mapping.

As shown in Fig. 7b, the difference in the number of purchased samples directly translates to the difference in $\phi(\mathcal{A})$. The average $\phi(\mathcal{A})$ of both mechanisms is monotonically increasing as B increases, but tends to grow at a lower rate. Moreover, there exists a significant improvement of BFreaMech, and the average $\phi(\mathcal{A})$ is 18.5%-22.3% higher than the baseline mechanism.

VII. Conclusion

In this paper, we designed an incentivized crowdsourcing system that acquires spectrum data periodically from users. The goal of our system is to maximize the average Kriging/prediction-error variance reduction $\phi(\mathcal{A})$ for a given budget. We first proposed a computationally efficient, individually rational and truthful incentive mechanism with a cardinality constraint. On top of it, we proposed a budget-feasible mechanism by translating the budget constraint to a suitable cardinality constraint with the bisection method.

We performed simulations to evaluate the performance of the proposed mechanisms, and compare them against a baseline mechanism. Our results show that the baseline mechanism



(a)

(b)

Fig. 7: Impact of budget B on the performance of BFeaMech and the baseline mechanism. (a) Average number of purchased samples as a function of B. (b) Average $\phi(\mathcal{A})$ as a function of B. BFeaMech is able to achieve an average $\phi(\mathcal{A})$ value 18.5%-22.3% higher than that of the baseline mechanism for different budgets.

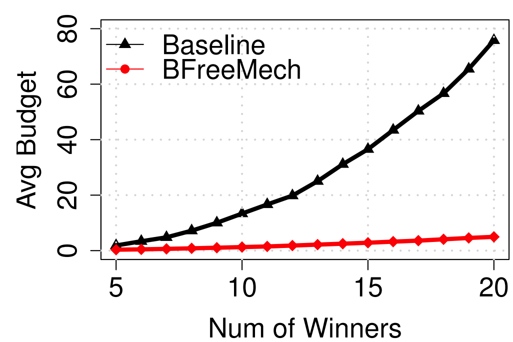


Fig. 8: Average budget as a function of number of winners k for the baseline mechanism and BFreeMech. For the same number of winners, the average budget required by the baseline mechanism is higher than the average total payment given by BFreeMech.

does not make full use of the budget in radio mapping, and the proposed mechanism achieves significantly better performance over the baseline mechanism, with a 18%-22% higher average $\phi(\mathcal{A})$ for different numbers of users and budgets.

Acknowledgments

We would like to thank Andrew Clark and Hossein Hosseini for their helpful discussions.

REFERENCES

[1] FCC, "Second report and order and memorandum opinion and order," 2008.

[2] C. Phillips, M. Ton, D. Sicker, and D. Grunwald, "Practical radio environment mapping with geostatistics," in *Dynamic Spectrum Access Networks (DYSPAN), 2012 IEEE International Symposium on*. IEEE, 2012, pp. 422–433.

[3] X. Ying, C. W. Kim, and S. Roy, "Revisiting tv coverage estimation with measurement-based statistical interpolation," in *Communication Systems and Networks (COMSNETS), 2015 Seventh International Conference on*. IEEE, 2015.

[4] A. Achtzehn, J. Riihijarvi, and P. Mahonen, "Improving accuracy for tvws geolocation databases: Results from measurement-driven estimation approaches," in *Dynamic Spectrum Access Networks (DYSPAN), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 392–403.

[5] J. Howe, "The rise of crowdsourcing," *Wired magazine*, vol. 14, no. 6, pp. 1–4, 2006.

[6] A. Nika, Z. Zhang, X. Zhou, B. Y. Zhao, and H. Zheng, "Towards commoditized real-time spectrum monitoring," in *Proceedings of the 1st ACM workshop on Hot topics in wireless*. ACM, 2014, pp. 25–30.

[7] O. Fatemieh, R. Chandra, and C. A. Gunter, "Secure collaborative sensing for crowd sourcing spectrum data in white space networks," in *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*. IEEE, 2010, pp. 1–12.

[8] D.-H. Shin, S. He, and J. Zhang, "Joint sensing task and subband allocation for large-scale spectrum profiling," in *The 34th IEEE Conference on Computer Communications (INFOCOM 2015)*, April-May 2015.

[9] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 173–184.

[10] Y. Singer, "Budget feasible mechanisms," in *Foundations of Computer Science (FOCS), 2010 51st Annual IEEE Symposium on*. IEEE, 2010, pp. 765–774.

[11] N. Chen, N. Gravin, and P. Lu, "On the approximability of budget feasible mechanisms," in *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*. SIAM, 2011, pp. 685–699.

[12] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1231–1239.

[13] L. G. Jaimes, I. Vergara-Laurens, and M. A. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 103–108.

[14] N. A. Cressie and N. A. Cassie, *Statistics for spatial data*. Wiley New York, 1993, vol. 900.

[15] P. J. R. Jr and P. J. Diggle, "geoR: a package for geostatistical analysis," *R-NEWS*, vol. 1, no. 2, pp. 14–18, June 2001, iSSN 1609-3631. [Online]. Available: http://CRAN.R-project.org/doc/Rnews/

[16] J. W. Van Groenigen, W. Siderius, and A. Stein, "Constrained optimisation of soil sampling for minimisation of the kriging variance," *Geoderma*, vol. 87, no. 3, pp. 239–259, 1999.

[17] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, 2008.

[18] A. Das and D. Kempe, "Algorithms for subset selection in linear regression," in *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM, 2008, pp. 45–54.

[19] A. Krause, H. B. McMahon, C. Guestrin, and A. Gupta, "Robust submodular observation selection," 2008.

[20] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functionsâĂŤi," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.

[21] V. Krishna, *Auction theory*. Academic press, 2009.

[22] R. B. Myerson, "Optimal auction design," *Mathematics of operations research*, vol. 6, no. 1, pp. 58–73, 1981.

[23] 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 *Sensor, Mesh and Ad Hoc Communications and Networks (SECON), 2012 9th Annual IEEE Communications Society Conference on*. IEEE, 2012, pp. 623–631.

[24] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013. [Online]. Available: http://www.R-project.org/