Frameworks for Privacy-Preserving Mobile Crowdsensing Incentive Mechanisms

Jian Lin, *Student Member, IEEE*, Dejun Yang, *Member, IEEE*, Ming Li, *Student Member, IEEE*, Jia Xu, *Member, IEEE*, and Guoliang Xue, *Fellow, IEEE*

Abstract—With the rapid growth of smartphones, mobile crowdsensing emerges as a new paradigm which takes advantage of the pervasive sensor-embedded smartphones to collect data efficiently. Many auction-based incentive mechanisms have been proposed to stimulate smartphone users to participate in the mobile crowdsensing applications and systems. However, none of them has taken into consideration both the bid privacy of smartphone users and the social cost. In this paper, we design two frameworks for privacy-preserving auction-based incentive mechanisms that also achieve approximate social cost minimization. In the former, each user submits a bid for a set of tasks it is willing to perform; in the latter, each user submits a bid for each task in its task set. Both frameworks select users based on platform-defined score functions. As examples, we propose two score functions, linear and log functions, to realize the two frameworks. We rigorously prove that both proposed frameworks achieve computational efficiency, individual rationality, truthfulness, differential privacy, and approximate social cost minimization. In addition, with log score function, the two frameworks are asymptotically optimal in terms of the social cost. Extensive simulations evaluate the performance of the two frameworks and demonstrate that our frameworks achieve bid-privacy preservation although sacrificing social cost.

Index Terms—Mobile crowdsensing, incentive mechanism, differential privacy

1 Introduction

Nowadays, the proliferation of smartphones is changing people's daily lives. With the advance of high-speed 3G/4G networks and more powerful embedded sensors (e.g., camera, accelerometer, compass, etc.), mobile crowdsensing emerges as a new paradigm which takes advantage of the pervasive sensor-embedded smartphones to collect data efficiently.

A typical mobile crowdsensing system consists of a cloud-based platform and a large number of smartphone users. The platform works as a sensing service buyer who posts the required sensing information and recruits a set of smartphone users to provide sensing services. Once selected by the platform, a smartphone user starts to collect the required data and sends it back to the platform. The potential effectiveness of mobile crowdsensing, especially with geographically distributed smartphone users, enables numerous mobile crowdsensing applications [34], [47], [53]. However, most of them assume that the smartphone users contribute to the platform voluntarily. In reality, smartphone users consume their own resources such as battery and sensing time while completing the sensing tasks. In addition, they might suffer from the potential privacy disclosure by sharing their sensed data with personal information (e.g., location tags and bid price). Therefore, smartphone users may be reluctant to participate in a mobile crowdsensing system and application, unless they are paid some rewards to compensate their resource consumption or potential privacy leaks. Since the number of participating smartphone users has a significant impact on the performance of the mobile crowdsensing systems, it is necessary to stimulate users to join the systems.

Auction is an efficient method to design incentive mechanisms. Many auction-based incentive mechanisms have been proposed for mobile crowdsensing [46], [47], [49], [51]. They are essentially reverse auctions in which the platform is the service buyer and the smartphone users are the bidders selling sensing services. In these mechanisms, the service buyer selects bidders according to their submitted task-bid pairs (elaborated in Section 3). The objectives of these mechanisms focus on either maximizing the total value gained by the platform or minimizing the total payment to the selected users. However, none of them takes users' privacy into consideration.

In most of the proposed truthful auction-based incentive mechanisms, bidders are stimulated to bid their true costs, which are private information of smartphone users. For transparency, the platform will publish the outcome of the auction mechanism, which consists of winning bidders and their payments. Ensuring transparency in the procurement

- *J. Lin and M. Li are with the Colorado School of Mines, Golden, CO 80401. E-mail: {jilin, mili}@mines.edu.*
- *D. Yang is with the Colorado School of Mines, Golden, CO 80401, and the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China. E-mail: djyang@mines.edu.*
- *J. Xu is with the Jiangsu Key Laboratory of Big Data Security and Intelligent Processing, Nanjing University of Posts and Telecommunications, Nanjing, Jiangsu 210023, China. E-mail: xujia@njupt.edu.cn.*
- *G. Xue is with Arizona State University, Tempe, AZ 85287. E-mail: xue@asu.edu.*

Manuscript received 28 Dec. 2016; revised 20 Nov. 2017; accepted 26 Nov. 2017. Date of publication 7 Dec. 2017; date of current version 29 June 2018. (Corresponding author: Dejun Yang.)
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TMC.2017.2780091

1536-1233 © 2017 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

procedure is essential to efficiency, as it enhances the competitiveness of public procurement [36]. Meanwhile, it has been proven by bid sale dealers for years that transparency leads to profit [37]. The FCC uses auctions to sell the licenses to transmit signals over specific bands of the electromagnetic spectrum, and releases the result of each auction online for transparency [3]. In recent years, many commercial platforms have put more emphasis on transparency as well, e.g., Auction.com [1], which is a trusted leader in the web-based real estate auction industry, and eBay [2], which is a multinational e-commerce corporation.

However, once the true cost of a smartphone user is reported to the platform, other bidders might infer this private information based on the published outcome. This is known as *inference attack* [20] (we give two examples in Section 3). Inference attack has been analyzed in many areas, e.g., multilevel secure databases [22], data mining [10], web-based applications [40] and mobile devices [31]. Protecting users' bid privacy is important because its disclosure might also incur threats to users' other private information, such as location [25],[44]. In this paper, we focus on designing truthful auction-based mechanisms to protect users' bid privacy.

To formalize the notion of users' bid privacy, we employ the concept of *differential privacy* [12]. Intuitively, a mechanism provides differential privacy if the change of one user's bid has limited impact on the outcome. We also leverage the *exponential mechanism* [33], a technique to design differentially private mechanisms, to preserve users' bid privacy.

In this paper, we study the problem of designing truthful mechanisms, which achieve computational efficiency, individual rationality, differential privacy, and approximate social cost minimization. We consider the scenario where there is one buyer and multiple sellers. Smartphone users act as bidders and submit their bids to compete for the chance of being selected to perform the corresponding tasks. Besides, smartphone users do not want others to know their own bid information. We first consider the single-bid model in which each user can only submit a set of tasks. Then we consider the multi-bid model in which each user can submit a bid for each task in its task set. For each of these two models, we propose a differentially private truthful auction-based framework, named BidGuard and Bid Guard-M, respectively. One important component of both frameworks is a platform-defined score function for selecting users. As examples, we propose two score functions to realize the frameworks.

The main contributions of this paper are as follows:

- In this paper, we propose two frameworks, BidGuard and BidGuard-M, for privacy-preserving mobile crowdsensing incentive mechanisms for two different models, which achieve computational efficiency, individual rationality, truthfulness, differential privacy, and approximate social cost minimization. Specifically, we design two different score functions, linear score function and log score function, to realize this two frameworks.
- With linear score function, BidGuard achieves $(\epsilon(e-1)/e, \delta)$-differential privacy and the social cost is at most $g\mathcal{OPT} + O(\ln n)$ with the probability of at least $1 - 1/n^{O(1)}$, where $\epsilon > 0$ is a constant, $\delta \in (0, \frac{1}{2}]$ and g is the cardinality of the largest user task set, e is the base of the natural logarithm, $\mathcal{OPT}$ is the optimal social cost, and n is the number of the users. BidGuard-M achieves $2m\epsilon$-differential privacy and the social cost is at most $\mathcal{OPT} + mO(\ln n)$ with the probability of at least $1 - 1/n^{O(1)}$, where m is the number of sensing tasks.
- With log score function, BidGuard achieves $(\epsilon(e-1)/e, \delta)$-differential privacy and the social cost is at most $2^t H_m \mathcal{OPT}$ with the probability of at least $1 - e^{-t}$ for any constant $t > 0$, where $H_m = \sum_{j=1}^{m} 1/j$, and m is the number of sensing tasks. BidGuard-M achieves $2m \log_{\frac{1}{2}}(\frac{1}{1+\Delta})\epsilon$-differential privacy and the social cost is at most $2^t\mathcal{OPT}$ with the probability of at least $1 - 1/n^{O(1)}$, where Δ is the maximum difference in the bidding price. In addition, both BidGuard and Bid Guard-M are proved to be asymptotically optimal.
- We evaluate the performance of BidGuard and BidGuard-M through simulations based on a real data set. Extensive numerical results demonstrate that both frameworks achieve bid-privacy preservation although sacrificing social cost.

The remainder of this paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, we introduce two system models and the objectives. In Sections 4 and 5, we present frameworks for the two models in detail and prove their properties, respectively. We evaluate the performance of our frameworks in Section 6. We conclude this paper in Section 7.

2 Related Work

In recent years, incentive mechanisms in mobile crowdsensing have been widely studied [16], [38]. As one of the pioneering works on designing incentive mechanisms for mobile crowdsensing, Yang et al. [48], [49] proposed two incentive mechanisms for both user-centric and platform-centric models using auction and Stackelberg game, respectively. The objectives of most of the state-of-art incentive mechanisms are either maximizing the total utility/value of the platform under a certain constraint (e.g., budget) [52] or minimizing the total payment of the platform [32]. Feng et al. [15] proposed a mechanism called TRAC, which takes into consideration the importance of location information when assigning sensing tasks.

Many pieces of works have explored the privacy-preserving mechanisms in mobile crowdsourcing. Most of them [17], [26] apply the spacial and temporal cloaking techniques like K-anonymity to blur users' locations in a cloaked area or cloaked time interval to preserve users' privacy. PEPSI [11] and AnonySense [39] focus on anonymous data collection, which could protect users' identities when they submit the tasks.

Some efforts have been specially made to protect users' privacy in mobile crowdsensing [8]. Although providing good performance in privacy preservation, the mechanisms in [14], [18], [27], [28], [29], [35], [43], [50] are based on cryptography techniques and do not take into consideration the users' strategic behaviors. Besides, all of the cryptography-based works are vulnerable to inference attack, since an attacker can infer users' private information through the published results. Sun et al. [41] proposed an auction-based

incentive mechanism which encrypts users' bids by oblivious transfer. But it does not solve the issue of inference attack because one user still can infer others' bids from the received payment. Jin et al. proposed a privacy-preserving approximately truthful incentive mechanism [23], which minimizes the total payment, and a privacy-preserving framework [24] for data aggregation. However, none of the above works has a performance guarantee on social cost. In this paper, our objectives are preserve users' bid privacy from inference attack while achieving approximate social cost minimization.

Differential privacy was first introduced by Dwork et al. [12]. The first differentially private auction mechanism was proposed by McSherry et al. [33]. They also incorporate exponential mechanism and mechanism design to achieve differential privacy with different objectives. General methods to design truthful mechanisms while still preserving differential privacy have been studied in [7], [21], [45]. However, our objective is different from above works. Recently, differential privacy has been used in other applications, e.g., location-based systems [4] and spatial crowdsouc-ing [42]. Zhu et al. [54] proposed the first differentially private spectrum auction mechanism, which achieves strategy-proofness and approximate revenue maximization. Note that our objective is to minimize the social cost, which differs from that in [54].

3 Models and Problem Formulation

In this section, we model the mobile crowdsensing system as a reverse auction and present two different models. Similar to most mobile crowdsensing systems [15], [47], [48], [49], [51], we consider a mobile crowdsensing system consisting of a platform and multiple smartphone users who are interested in performing sensing tasks. In the first model, each user can submit only one task-bid pair. Our second model allows each user to submit multiple task-bid pairs and can be assigned to work on multiple tasks. Then we describe the threat models, which threaten both of the models. At the end of this section, we present some important properties and give our design objective.

3.1 Single-Bid Model

The platform first publicizes a set $\mathcal{T} = \{t_1, t_2, \ldots, t_m\}$ of m sensing tasks. Assume there is a set $\mathcal{U} = \{1, 2, \ldots, n\}$ of $n \geqslant 2$ smartphone users. Each user i has a task set $\Gamma_i \subseteq \mathcal{T}$, which it can perform. Each Γ_i is associated with a cost c_i, which is a private information of user i. The platform selects a subset of users $\mathcal{S} \subseteq \mathcal{U}$ to complete all the sensing tasks in $\mathcal{T}$. At last, the platform calculates the payment p_i for each selected user $i \in \mathcal{S}$. Let $\overrightarrow{p} = (p_1, p_2, \ldots, p_n)$ denote the payment profile. The utility of any user $i \in \mathcal{U}$ is

$$u_i = \begin{cases} p_i - c_i, & \text{if } i \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

In this paper, we model the interactive process between the platform and the users as a sealed-bid reverse auction, where the platform buys sensing service and the users are bidders who sell sensing service. In order to prevent the monopoly and guarantee the quality of sensing task, we assume each task in $\mathcal{T}$ can be completed by more than one user in $\mathcal{U}$. This assumption is reasonable for mobile crowdsensing as made in [15]. If a task in $\mathcal{T}$ can only be completed by at most one user in $\mathcal{U}$, we simply remove it from $\mathcal{T}$.

At the beginning of this auction, each user $i \in \mathcal{U}$ submits a task-bid pair $\beta_i = (\Gamma_i, b_i)$ to the platform, where b_i is user i's bid, representing the minimum price user i wants to sell its sensing service for. Note that in a truthful auction-based incentive mechanism, users are stimulated to bid their true costs, i.e., $b_i = c_i$. Without loss of generality, we assume that each user's bid is bounded by $[b_{min}, b_{max}]$, where b_{min} is normalized to 1 and b_{max} is a constant. Let Δ denote the difference between b_{max} and b_{min}. Let $\overrightarrow{\beta} = (\beta_1, \beta_2, \ldots, \beta_n)$ denote the task-bid profile. Given the task-bid profile $\overrightarrow{\beta}$, the platform determines the outcome of the auction, which consists of selected winning users $\mathcal{S}$ and the payment profile $\overrightarrow{p}$.

3.2 Multi-Bid Model

In the single-bid model, each user submit a bid for a set of tasks. In the multi-bid model, each user is allowed to submit a bid for each task in its task set, and each user can be assigned to work on multiple tasks.

The definitions of $\mathcal{T}, \mathcal{U}, \mathcal{S}, \Gamma_i, \overrightarrow{p}$ and Δ are the same as in Section 3.1. In the multi-bid model, for each user $i \in \mathcal{U}$, each task t_i^k in Γ_i has an associated cost c_i^k. Each user i submits a set $\mathcal{B}_i = \{\beta_i^1, \beta_i^2, \ldots, \beta_i^{k_i}\}$ of $k_i = |\Gamma_i|$ task-bid pairs. Each task-bid pair is denoted by $\beta_i^k = (t_i^k, b_i^k)$, where t_i^k is a single task from Γ_i, and b_i^k is the minimum price user i wants to sell its sensing service for t_i^k. Note that in a truthful auction-based incentive mechanism, users are stimulated to bid their true costs, i.e., $b_i^k = c_i^k$. Let $\overrightarrow{\mathcal{B}} = (\mathcal{B}_1, \mathcal{B}_2, \ldots, \mathcal{B}_n)$ denote the task-bid profile. Given the task-bid profile $\overrightarrow{\mathcal{B}}$, the platform determines the winning task-bid pair set $\mathcal{B}_W \subseteq \bigcup_{i \in \mathcal{U}} \mathcal{B}_i$ such that $\bigcup_{\beta_i^k \in \mathcal{B}_W} t_i^k = \mathcal{T}$. For each winning task-bid pair $\beta_i^k \in \mathcal{B}_W$, the platform calculates a payment p_i^k. A user i is called a winner and be added into $\mathcal{S}$ if it has at least one winning task-bid pair, i.e., $\mathcal{B}_i \cap \mathcal{B}_W \neq \emptyset$. The payment for each winner i is $p_i = \sum_{\beta_i^k \in \mathcal{B}_i \cap \mathcal{B}_W} p_i^k$. The utility of any user $i \in \mathcal{U}$ is

$$u_i = \begin{cases} p_i - \sum_{\beta_i^k \in \mathcal{B}_i \cap \mathcal{B}_W} c_i^k, & \text{if } i \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

3.3 Threat Models

Threats to Incentive. We assume that users are selfish but rational. Hence user i could report a bid b_i differs from its true cost c_i, i.e., $b_i \neq c_i$ in the single-bid model or report a bid $b_i^k \neq c_i^k$ in the multi-bid model to maximize its own utility. We also assume that user i does not misreport its task set Γ_i in the single-bid model as in [15], [47], [48], [49], [51], and does not misreport any $t_i^k \in \Gamma_i$ in the multi-bid model.[1] Other threats to incentive (e.g., collusion among bidders) are out of the scope of this paper.

Threats to Privacy. As mentioned earlier, bidders are stimulated to bid their true costs in a truthful auction-based incentive mechanism, i.e., $b_i = c_i$ in the single-bid model

1. In the single-bid model, if user i reports Γ_i' containing tasks not in Γ_i, i.e., $\Gamma_i' \setminus \Gamma_i \neq \emptyset$, it cannot finish Γ_i' when selected. If user i reports $\Gamma_i' \subset \Gamma_i$ with c_i, the probability of user i being selected will not increase according to our mechanism. The case where user i misreports both Γ_i and c_i is challenging, because calculating the true cost of $\Gamma_i' \subset \Gamma_i$ is still an open question. In the multi-bid model, if user i reports β_i^k containing tasks not in Γ_i, i.e., $t_i'^k \notin \Gamma_i$, it cannot finish $t_i'^k$ when selected.

TABLE 1
Example Showing the Inference Attack in Single-Bid Model

β_i \ User	1	2	3	4	5
Γ_i	t_1, t_2	t_1	t_1, t_3	t_1, t_2	t_1, t_3
b_i	\$3	\$1	\$4	\$5	\$5

TABLE 2
Example Showing the Inference Attack in Multi-Bid Model

β_i^k \ User	1		2	3		4		5	
t_i^k	t_1	t_2	t_1	t_1	t_3	t_1	t_2	t_1	t_3
b_i^k	\$1.5	\$1.5	\$1	\$1.6	\$2.4	\$3	\$2	\$2.5	\$2.5

and $b_i^k = c_i^k$ in the multi-bid model. However, one bidder could infer other bidders' bid according to the outcome of the mechanism. This inference attack can be seen from the following examples.

We first consider the single-bid model, suppose there are 5 users in the system and their task-bid pairs $\beta_i = (\Gamma_i, b_i)$, $i \in [1, 5]$ are shown in Table 1. The platform publicizes a set of 3 sensing tasks $\mathcal{T} = \{t_1, t_2, t_3\}$. According to the proposed truthful mechanism in TRAC [15], the winning users $\mathcal{S} = \{2, 1, 3\}$. Suppose $user$ 5 is a bidder who want to infer other bidders' bid, and it changes its bid b_5 from \$5 to \$3 in the next auction while the other four bidders do not change their task-bid pairs. The winning users of the new auction is $\mathcal{S} = \{2, 1, 5\}$. Since the platform publishes the outcome of the mechanism for transparency, $user$ 5 could know the results and infer that $user$ 3's bid is between \$3 and \$5 by the fact that if it bids \$5 it will be replaced by $user$ 3 and if it bid \$3 it will replace $user$ 3. We can see that, after many rounds of auction, $user$ 5 might narrow down $user$ 3's bid range, and even infer the exact value in some cases.

Next we consider the inference attack in multi-bid model using the example shown in Table 2. According to the proposed truthful mechanism in TRAC [15], the winning bid-pairs are $\mathcal{B}_\mathcal{W} = \{\beta_2^1, \beta_1^2, \beta_3^2\}$, and thus $\mathcal{S} = \{2, 1, 3\}$. Suppose $user$ 5 is a bidder who want to infer other bidders' bid, and it changes its bid b_5^2 from \$2.5 to \$2 in the next auction while the other four bidders do not change their task-bid pairs. Then the winning bid-pairs of the new auction are $\mathcal{B}_\mathcal{W} = \{\beta_2^1, \beta_1^2, \beta_5^2\}$. Based on this outcome, $user$ 5 could infer that $user$ 3's bid for t_3 is between \$2 and \$2.5 by the fact that if it bids \$2.5 it will be replaced by $user$ 3 and if it bids \$2 it will replace $user$ 3. After many rounds of auction, $user$ 5 might also narrow down $user$ 3's bid range, and even infer the exact value in some cases.

This inference attack is practical in most mobile crowdsensing applications, e.g., [34], [53], where tasks are publicized periodically for collecting dynamic sensing data. Protecting users' bid privacy from such inference attack is important because its disclosure might also incur threats to users' other private information, such as location [25], [44]. For example, in [25] each user i's cost of task c_i is modeled as a linear function of its distance d_i to the task. In a truthful mechanism, user i's bid $b_i = c_i$. Therefore, an attacker can infer user i's location inside a suspicion region, which is the circle centered at the task with radius d_i, by inferring its bid b_i. Besides, an attacker can also improve the inference accuracy by narrowing down the victim's bid through many rounds of auction.

3.4 Desired Properties

We consider the following important properties.

- *Computational Efficiency:* A mechanism is computationally efficient if it terminates in polynomial time.
- *Individual Rationality:* A mechanism is individually rational if each user will have a non-negative utility when bidding its true cost.
- *Truthfulness:* A mechanism is truthful if any user's utility is maximized when bidding its true cost.
- *Social Cost Minimization:* A mechanism achieves social cost minimization if the total cost of the users in $\mathcal{S}$ is minimized subject to certain constraints on $\mathcal{S}$.

In addition, we consider users' bid privacy preservation.

Definition 1 (Differential Privacy [12]). *A randomized function M has ϵ-differential privacy if for any two input sets A and B with a single input difference, and for any set of outcomes $\mathcal{O} \subseteq Range(M)$*

$$Pr[M(A) \in \mathcal{O}] \leq \exp(\epsilon) \times Pr[M(B) \in \mathcal{O}].$$

In this paper, the randomized function M is corresponding to our frameworks, and $Range(M)$ is the outcome space of the frameworks. One relaxation of differential privacy is as follows.

Definition 2 (Approximate Differential Privacy [13]). *A randomized function M gives (ϵ, δ)-differential privacy if for any two input sets A and B with a single data difference, and for any set of outcomes $\mathcal{O} \subseteq Range(M)$*

$$Pr[M(A) \in \mathcal{O}] \leq \exp(\epsilon) \times Pr[M(B) \in \mathcal{O}] + \delta.$$

The truthfulness of an auction mechanism is guaranteed by the following theorem.

Theorem 1 ([5]). *Let $Pr_i(z)$ denote the probability that bidder i is selected when its bid is z. A mechanism with bids $\overrightarrow{b}$ and payments $\overrightarrow{p}$ is truthful in expectation if and only if, for any bidder i,*

1) *$Pr_i(z)$ is monotonically non-increasing in b_i;*
2) *$\int_0^\infty Pr_i(z)dz < \infty$;*
3) *The expected payment satisfies $E[p_i] = b_i Pr_i(b_i) + \int_{b_i}^\infty Pr_i(z)dz$.*

Next, we introduce the concept of the exponential mechanism and its properties. In the literature of differential privacy, the exponential mechanism is often used to design privacy-preserving mechanisms. A key component of the exponential mechanism is the score function $f(A, o)$, which maps the input set A and an outcome $o \in \mathcal{O}$ to a real-valued score. The score represents how good the outcome o is for the input set A compared with the optimal outcome.

Exponential Mechanism $\epsilon_f^\epsilon(A)$. Given an outcome space $\mathcal{O}$, an input set A, a score function f and a small constant ϵ, the exponential mechanism $\epsilon_f^\epsilon(A)$ chooses an outcome $o \in \mathcal{O}$ with probability

$$Pr\left[\epsilon_f^\epsilon(A) = o\right] \propto \exp(\epsilon f(A, o)).$$

Let Λ denote an upper-bound of the difference of any two input sets, the exponential mechanism has the following properties.

Theorem 2 ([33]). *The exponential mechanism gives $2\epsilon\Lambda$-differential privacy.*

Theorem 3 ([19]). *For any $\alpha \geq 0$, the exponential mechanism, when used to select an output $o \in \mathcal{O}$, $\epsilon_f^{\epsilon}(A)$ yields $2\epsilon\Lambda$-differential privacy, letting $\mathcal{O}^*$ be the subset of $\mathcal{O}$ achieving $f(A,o) = \max_o f(A,o)$, ensures that*

$$Pr\left[f(A,\epsilon_f^{\epsilon}(A)) < \max_o f(A,o) - \ln(|\mathcal{O}|/|\mathcal{O}^*|)/\epsilon - \alpha/\epsilon\right] \leq \exp(-\alpha).$$

3.5 Design Objective

The goal of our framework design is to minimize the social cost while achieving computational efficiency, individual rationality, truthfulness and differential privacy. Specifically, the minimization problem in the single-bid model is referred to as the Social Cost Minimization (SCM) problem and the minimization problem in the multi-bid model is referred to as the SCM-M problem. Next, we give the formal formulation of the SCM problem and the SCM-M problem, respectively.

SCM Problem. Given a task set $\mathcal{T}$ and a user set $\mathcal{U}$, the goal of the SCM-S problem is to find a subset of users $\mathcal{S} \subseteq \mathcal{U}$, such that $C(\mathcal{S}) = \sum_{i\in\mathcal{S}} c_i$ is minimized subject to $\bigcup_{i\in\mathcal{S}} \Gamma_i = \mathcal{T}$.

SCM-M Problem. Given a task set $\mathcal{T}$ and a user set $\mathcal{U}$, the goal of the SCM-M problem is to find a subset of users $\mathcal{S} \subseteq \mathcal{U}$ and their assigned task-bid pairs $\mathcal{B}_W$, such that $C(\mathcal{B}_W) = \sum_{\beta_i^k \in \mathcal{B}_W} c_i^k$ is minimized subject to $\bigcup_{\beta_i^k \in \mathcal{B}_W} t_i^k = \mathcal{T}$.

Note that SCM problem is challenging because it is NP-hard (proved by Theorem 4 in [30]), let alone combining with computational efficiency, individual rationality, truthfulness and differential privacy. Although SCM-M can be solved optimally, it is still challenging when combining with the other properties. Therefore, we aim to design differentially private truthful frameworks with theoretically guaranteed approximate social cost.

4 BidGuard: Differentially Private Auction Framework for the Single-bid Model

In this section, we design and analyze BidGuard, a differentially private auction framework for the Single-bid Model.

4.1 Design Rationale

BidGuard integrates the exponential mechanism with the reverse auction to achieve computational efficiency, individual rationality, truthfulness, differential privacy and approximate social cost minimization. In this framework, users are selected iteratively. In each iteration, redundant users are eliminated and each remaining user is assigned a probability to be selected. The framework then selects one of them as the winner based on the probability distribution. Specifically, the probability of a user to be selected is set according to a specific criterion. The above processes repeats until all the sensing tasks can be completed by the selected users. Finally, the framework computes the payment to each winner.

4.2 Design of BidGuard

In this section, we will describe BidGuard in detail. As illustrated in Algorithm 1, BidGuard consists of three phases: user screening, winner selection, and payment determination. It executes these three phases iteratively until all the sensing tasks can be completed by the selected users.

Algorithm 1. BidGuard

Input: A set of sensing tasks $\mathcal{T}$, a set of users $\mathcal{U}$, submitted task-bid profile $\overrightarrow{\beta}$, and differential privacy parameters $\epsilon > 0$ and $\delta \in (0, \frac{1}{2}]$.
Output: A set of winners $\mathcal{S}$ and a payment profile $\overrightarrow{p}$.

$\mathcal{S} \leftarrow \emptyset, \mathcal{T}_c \leftarrow \emptyset, \mathcal{R} \leftarrow \mathcal{U}$;
foreach $i \in \mathcal{U}$ **do** $p_i \leftarrow 0$
while $\mathcal{T}_c \neq \mathcal{T}$ **do**
  foreach $i \in \mathcal{R}$ **do**
    if $\Gamma_i \subseteq \mathcal{T}_c$ **then** $\mathcal{R} \leftarrow \mathcal{R} \setminus \{i\}$
  end
  foreach $i \in \mathcal{R}$ **do**
    Calculate the probability $Pr_i(b_i)$ of each user being selected according to the score function;
  end
  Select one user randomly, denoted by i', according to the computed probability distribution;
  $\mathcal{S} \leftarrow \mathcal{S} \cup \{i'\}, \mathcal{T}_c \leftarrow \mathcal{T}_c \cup \Gamma_{i'}, \mathcal{R} \leftarrow \mathcal{R} \setminus \{i'\}$;
end
foreach $i \in \mathcal{S}$ **do** $p_i \leftarrow b_i + \frac{\int_{b_i}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i)}$
return $\mathcal{S}$ and $\overrightarrow{p}$.

1) User Screening Phase. BidGuard will eliminate all the redundant users, whose task set can be completed by the currently selected users. The set of remaining users is denoted by $\mathcal{R}$.

2) Winner Selection Phase. BidGuard will assign each user $i \in \mathcal{R}$ a probability of being selected as follows. It first computes a criterion $r(\beta_i)$, which is the bid divided by the number of tasks that cannot be completed by the currently selected users, i.e.,

$$r(\beta_i) = \frac{b_i}{|\Gamma_i - \mathcal{T}_c|}, \tag{1}$$

where $\mathcal{T}_c$ is the set of tasks that can be completed by the currently selected users. BidGuard selects the user with the lowest $r(\beta_i)$ in each iteration. To apply the exponential mechanism, we need to design a score function, which is a non-increasing function of $r(\beta_i)$. The probability of each user to be selected is set according to the value of the score function.

3) Payment Determination Phase. Let $Pr_i(z)$ denote the probability of user i being selected with bid z. According to Theorem 1, the payment to winner i is

$$p_i = b_i + \frac{\int_{b_i}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i)}.$$

4.3 Design of Score Functions

To apply the exponential mechanism, we need to design a score function. Specifically, we design two score functions, *linear score function* and *log score function*. We will show that they have different theoretical bounds on the social cost (Section 4.4) and performance in simulations (Section 6).

Linear Score Function. $f_{LIN}(x) = 1 - x$. For any bidder $i \in \mathcal{R}$, the probability to be selected in each iteration is

$$Pr_i(b_i) \propto \begin{cases} \exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big), & \text{if } i \in \mathcal{R}; \\ 0, & \text{otherwise,} \end{cases}$$

where $\epsilon' = \epsilon/(e\Delta\ln(e/\delta))$. Note that in order to guarantee the value of the score function is nonnegative, we normalize $r(\beta_i)$, i.e., $\frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|}$. Then the probability is

$$Pr_i(b_i) = \begin{cases} \frac{\exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}{\sum_{j \in \mathcal{R}} \exp\Big(\epsilon'(1 - \frac{b_j}{b_{max}|\Gamma_j - \mathcal{T}_c|})\Big)}, & \text{if } i \in \mathcal{R}; \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

Log Score Function. $f_{LOG}(x) = \log_{1/2} x$. For any bidder $i \in \mathcal{R}$, the probability to be selected in each iteration is

$$Pr_i(b_i) \propto \begin{cases} \exp\Big(\epsilon' \log_{1/2} \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|}\Big), & \text{if } i \in \mathcal{R}; \\ 0, & \text{otherwise,} \end{cases}$$

where $\epsilon' = \epsilon/(e\ln(e/\delta)\log_{1/2}(1/(1+\Delta)))$. We also normalize the $r(\beta_i)$, i.e., $\frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|}$ to guarantee the value of the score function is nonnegative. Then the probability is

$$Pr_i(b_i) = \begin{cases} \frac{\exp\Big(\epsilon' \log_{1/2} \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|}\Big)}{\sum_{j \in \mathcal{R}} \exp\Big(\epsilon' \log_{1/2} \frac{b_j}{b_{max}|\Gamma_j - \mathcal{T}_c|}\Big)}, & \text{if } i \in \mathcal{R}; \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Throughout the rest of this paper, we denote the BidGuard with linear score function f_{LIN} and log score function f_{LOG} by LIN and LOG, respectively.

Illustrating Example. We use the example in Table 1 to illustrate how LIN works. Assuming $b_{min} = 1$, $b_{max} = 6$, then $\Delta = 5$. Let the differential privacy parameters $\epsilon = 0.1$ and $\delta = 0.5$, then $\epsilon' = 0.1/(e \times 5\ln(e/0.5))$. At the beginning, $\mathcal{T} = \{t_1, t_2, t_3\}$, $\mathcal{S} = \emptyset$, $\mathcal{T}_c = \emptyset$, $\mathcal{R} = \mathcal{U} = \{1, 2, 3, 4, 5\}$. LIN starts to select users iteratively. In the first iteration, LIN calculates $|\Gamma_i - \mathcal{T}_c|$ for each user $i \in \mathcal{R}$. We have $|\Gamma_1 - \mathcal{T}_c| = 2$, $|\Gamma_2 - \mathcal{T}_c| = 1$, $|\Gamma_3 - \mathcal{T}_c| = 2$, $|\Gamma_4 - \mathcal{T}_c| = 2$, and $|\Gamma_5 - \mathcal{T}_c| = 2$. Based on (2), LIN calculates the probability of every user in $\mathcal{R}$ to be selected in this iteration, e.g., $Pr_1(3) = \exp(0.7\epsilon')/(\exp(0.7\epsilon') + \exp(0.8\epsilon') + \exp(0.6\epsilon') + \exp(0.5\epsilon') + \exp(0.5\epsilon'))$. LIN selects a user based on the calculated probability distribution. Assume LIN selects *user* 1, then $\mathcal{S} = \{1\}$, $\mathcal{T}_c = \{t_1, t_2\}$, $\mathcal{R} = \{3, 5\}$. At the beginning of the second iteration, LIN calculates $|\Gamma_i - \mathcal{T}_c|$ for the remaining users *user* 3 and *user* 5. We have $|\Gamma_3 - \mathcal{T}_c| = 1$ and $|\Gamma_5 - \mathcal{T}_c| = 1$. Then LIN calculates the probabilities of *user* 3 and *user* 5 to be selected in this iteration according to (2). We have $Pr_3(4) = \exp(0.2\epsilon')/(\exp(0.2\epsilon') + \exp(\epsilon'))$ and $Pr_5(5) = \exp(\epsilon')/(\exp(0.2\epsilon') + \exp(\epsilon'))$. Assume *user* 3 is selected in this iteration, then LIN terminates since $\mathcal{T}_c = \mathcal{T}$. At last, LIN calculates the payment to all the selected users, i.e., *user* 1 and *user* 3. We have $p_1 = 3 + \frac{\int_3^6 Pr_1(z)dz}{Pr_1(3)}$ and $p_3 = 4 + \frac{\int_4^6 Pr_3(z)dz}{Pr_3(4)}$.

4.4 Analysis of BidGuard

In this section, we first analyze the properties of LIN.

Theorem 4. LIN *achieves computational efficiency, individual rationality, truthfulness, and* $(\epsilon(e-1)/e, \delta)$*-differential privacy, where* $\epsilon > 0$ *and* $\delta \in (0, \frac{1}{2}]$ *are constants,* e *is the base of the natural logarithm. In addition, it has social cost at most* $g\mathcal{OPT} + O(\ln n)$ *with probability at least* $1 - 1/n^{O(1)}$*, where* g *is the cardinality of the largest user task set,* $\mathcal{OPT}$ *is the optimal social cost of the SCM problem, and* n *is the number of users.*

Proof. We first prove the computational efficiency. The outer while-loop (Lines 3-12) will run at most m iterations since there are m tasks. Meanwhile, the two inner for-loops (Lines 4-6) and (Lines 7-9) will run at most n iterations since there are n users. Therefore, the total computational complexity of LIN is $O(mn)$. The individual rationality is guaranteed by the fact that the payment to each winner i is $p_i = b_i + \frac{\int_{b_i}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i)} \geq b_i$. In order to prove the rest of this theorem, we prove the following lemmas. □

Lemma 1. LIN *is truthful.*

Proof. According to (2) and (3), the probability $Pr_i(b_i)$ of user i being selected in BidGuard is monotonically non-increasing in its bid b_i. In addition, no bid is greater than b_{max} in our model. Thus we have $\int_0^\infty Pr_i(z)dz = \int_0^{b_{max}} Pr_i(z)dz < \infty$. Furthermore, we have

$$\begin{aligned} &E[p_i] \\ &= (1 - Pr_i(b_i)) \times 0 + Pr_i(b_i) \times \left(b_i + \frac{\int_{b_i}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i)}\right) \\ &= b_i Pr_i(b_i) + \int_{b_i}^{\infty} Pr_i(z)dz. \end{aligned}$$

Then, according to Theorem 1, the lemma holds. □

Lemma 2. *For any constants* $\epsilon > 0$ *and* $\delta \in (0, \frac{1}{2}]$*,* LIN *achieves* $(\epsilon(e-1)/e, \delta)$*-differential privacy, where* e *is the base of the natural logarithm.*

Proof. Let $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$ be two input task-bid profiles that differ in any user d's bid, respectively. Let $M(\overrightarrow{\beta})$ and $M(\overrightarrow{\beta'})$ denote the sequences of users selected by LIN with inputs $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$, respectively. We show that LIN, even revealing the order in which the users are chosen, achieves differential privacy for an arbitrary sequence of users $\mathbb{I} = i_1, i_2, \ldots, i_l$ of arbitrary length l. We consider the relative probability of LIN for given task-bid inputs $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$

$$\begin{aligned} \frac{Pr\Big[M(\overrightarrow{\beta}) = \mathbb{I}\Big]}{Pr\Big[M(\overrightarrow{\beta'}) = \mathbb{I}\Big]} &= \prod_{j=1}^{l} \frac{\frac{\exp\Big(\epsilon'(1 - \frac{b_{i_j}}{b_{max}|\Gamma_{i_j} - \mathcal{T}_c|})\Big)}{\sum_{i \in \mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}}{\frac{\exp\Big(\epsilon'(1 - \frac{b'_{i_j}}{b_{max}|\Gamma_{i_j} - \mathcal{T}_c|})\Big)}{\sum_{i \in \mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b'_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}} \\ &= \prod_{j=1}^{l} \frac{\exp\Big(\epsilon'(1 - \frac{b_{i_j}}{b_{max}|\Gamma_{i_j} - \mathcal{T}_c|})\Big)}{\exp\Big(\epsilon'(1 - \frac{b'_{i_j}}{b_{max}|\Gamma_{i_j} - \mathcal{T}_c|})\Big)} \\ &\quad \times \prod_{j=1}^{l} \frac{\sum_{i \in \mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b'_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}{\sum_{i \in \mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}, \end{aligned}$$

where $\mathcal{U}_j = \mathcal{U} \setminus \{i_1, i_2, \ldots, i_{j-1}\}$ and the first equation is based on (2). We then prove this lemma by cases. When $b_d < b'_d$, the second product is at most 1 because the factor for any $j \in [1, l]$ is less than 1 if $d \in \mathcal{U}_j$ and equal to 1 otherwise. Therefore, we have

$$\begin{aligned}\frac{Pr\Big[M(\overrightarrow{\beta}) = i_1, i_2, \ldots, i_l\Big]}{Pr\Big[M(\overrightarrow{\beta'}) = i_1, i_2, \ldots, i_l\Big]} &\leq \frac{\exp\Big(\epsilon'(1 - \frac{b_d}{b_{max}|\Gamma_d - \mathcal{T}_c|})\Big)}{\exp\Big(\epsilon'(1 - \frac{b'_d}{b_{max}|\Gamma_d - \mathcal{T}_c|})\Big)} \\ &= \exp\left(\epsilon' \frac{b'_d - b_d}{b_{max}|\Gamma_d - \mathcal{T}_c|}\right) \\ &\leqslant \exp\big(\epsilon'(b'_d - b_d)\big) \\ &\leqslant \exp(\epsilon'\Delta).\end{aligned}$$

When $b_d \geq b'_d$, the first product is at most 1 because the factor for any $j \in [1, l]$ is less than 1 if $i_j = d$ and equal to 1 otherwise. In the remainder of the proof, we focus on this case. Therefore, we have

$$\begin{aligned}&\frac{Pr\Big[M(\overrightarrow{\beta}) = i_1, i_2, \ldots, i_l\Big]}{Pr\Big[M(\overrightarrow{\beta'}) = i_1, i_2, \ldots, i_l\Big]} \\ &\leqslant \prod_{j=1}^{l} \frac{\sum_{i\in\mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b'_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}{\sum_{i\in\mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)} \\ &= \prod_{j=1}^{l} \frac{\sum_{i\in\mathcal{U}_j} \exp\Big(\epsilon' \frac{\theta_i}{|\Gamma_i - \mathcal{T}_c|}\Big)\exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)}{\sum_{i\in\mathcal{U}_j} \exp\Big(\epsilon'(1 - \frac{b_i}{b_{max}|\Gamma_i - \mathcal{T}_c|})\Big)} \\ &= \prod_{j=1}^{l} E_{i\in\mathcal{U}_j}\left[\exp\left(\epsilon' \frac{\theta_i}{|\Gamma_i - \mathcal{T}_c|}\right)\right] \\ &\leqslant \prod_{j=1}^{l} E_{i\in\mathcal{U}_j}[\exp(\epsilon'\theta_i)],\end{aligned}$$

where $\theta_i = b'_i - b_i$. For all $x \leqslant 1$, $e^x \leqslant 1 + (e-1)\cdot x$. Therefore, for all $\epsilon' \leqslant 1$, we have

$$\begin{aligned}\prod_{j=1}^{l} E_{i\in\mathcal{U}_j}[\exp(\epsilon'\theta_i)] &\leqslant \prod_{j=1}^{l} E_{i\in\mathcal{U}_j}[1 + (e-1)\epsilon'\theta_i] \\ &\leqslant \exp\left((e-1)\epsilon' \sum_{j=1}^{l} E_{i\in\mathcal{U}_j}[\theta_i]\right).\end{aligned}$$

Lemma B.2 in [19] implies that $Pr[\sum_{j=1}^{l} E_{i\in\mathcal{U}_j}[\theta_i] > \Delta\ln(e/\delta)] \leqslant \delta$. Let $\mathcal{O}$ denote the outcome space, where each $o \in \mathcal{O}$ is a sequence of users $i_1, i_2, \ldots, i_l$. We split $\mathcal{O}$ into two sets $\mathcal{O}'$ and $\mathcal{O}''$, where $\mathcal{O}' = \{o \in \mathcal{O} | \sum_{j=1}^{l} E_{i\in\mathcal{U}_j}[\theta_i] \leq \Delta\ln(e/\delta)\}$ and $\mathcal{O}'' = \mathcal{O} \setminus \mathcal{O}'$. Thus we have

$$\begin{aligned}&Pr\Big[M(\overrightarrow{\beta}) \in \mathcal{O}\Big] \\ &= \sum_{o\in\mathcal{O}} Pr\Big[M(\overrightarrow{\beta}) = o\Big] \\ &= \sum_{o\in\mathcal{O}'} Pr\Big[M(\overrightarrow{\beta}) = o\Big] + \sum_{o\in\mathcal{O}''} Pr\Big[M(\overrightarrow{\beta}) = o\Big] \\ &\leq \sum_{o\in\mathcal{O}'} \exp((e-1)\epsilon'\Delta\ln(e/\delta)) Pr\Big[M(\overrightarrow{\beta'}) = o\Big] + \delta \\ &\leq \exp((e-1)\epsilon'\Delta\ln(e/\delta) Pr\Big[M(\overrightarrow{\beta'}) \in \mathcal{O}\Big] + \delta \\ &= \exp(\epsilon(e-1)/e) Pr\Big[M(\overrightarrow{\beta'}) \in \mathcal{O}\Big] + \delta.\end{aligned}$$

The lemma holds. □

Lemma 3. *With probability at least $1 - 1/n^{O(1)}$, LIN has social cost at most $g\mathcal{OPT} + O(\ln n)$, where g is the cardinality of the largest user task set, $\mathcal{OPT}$ is the optimal social cost of the SCM problem, and n is the number of users.*

Proof. Let $\mathcal{S}^*$ denote the optimal solution to the SCM problem. For LIN, we consider a sequence $\mathcal{W}$ of winners according to the order they are selected, i.e., $\mathcal{W} = w_1, w_2, \ldots, w_l$.

For each $w_i, 1 \leq i \leq l$, let $\mathcal{W}_i$ denote the set of users satisfying $\forall j \in \mathcal{W}_i$:

1) $j \in \mathcal{S}^*$;
2) $\Gamma_j \cap \Gamma_{w_i} \neq \emptyset$;
3) $\Gamma_j \cap \Gamma_{w_k} = \emptyset, \forall k \in [1, i-1]$;

$\mathcal{W}_i$ is the set of users in $\mathcal{S}^*$ but not in $\mathcal{W}$ because of w_i. For truthful mechanisms, we have $b_i = c_i$. According to Theorem 3, by taking $\alpha = O(\ln n)$, we have

$$1 - \frac{c_{w_i}}{|\Gamma_{w_i} - \mathcal{T}_c|} \geq 1 - \frac{c_j}{|\Gamma_j - \mathcal{T}_c|} - O(\ln n),$$

with a probability of at least $1 - 1/n^{O(1)}$. This implies that

$$c_j \geq \frac{c_{w_i}}{|\Gamma_{w_i} - \mathcal{T}_c|} \cdot |\Gamma_j - \mathcal{T}_c| - O(\ln n),$$

with a probability of at least $1 - 1/n^{O(1)}$.
Summing over all $j \in \mathcal{W}_i$, we have

$$\begin{aligned}\sum_{j\in\mathcal{W}_i} c_j &\geq \left(\frac{c_{w_i}}{|\Gamma_{w_i} - \mathcal{T}_c|} - O(\ln n)\right) \cdot \sum_{j\in\mathcal{W}_i} |\Gamma_j - \mathcal{T}_c| \\ &\geq \frac{c_{w_i}}{|\Gamma_{w_i} - \mathcal{T}_c|} - O(\ln n),\end{aligned}$$

with a probability of at least $1 - 1/n^{O(1)}$. The first inequality holds because $\sum_{j\in\mathcal{W}_i} |\Gamma_j - \mathcal{T}_c| \geq |\mathcal{W}_i|$. The second inequality holds because $\sum_{j\in\mathcal{W}_i} |\Gamma_j - \mathcal{T}_c| \geq 1$. Note that $|\Gamma_{w_i} - \mathcal{T}_c|$ can be upper bounded by a constant g, which is the cardinality of the largest user task set. Therefore, we have

$$\sum_{j\in\mathcal{W}_i} c_j \geq \frac{c_{w_i}}{g} - O(\ln n).$$

Summing over all $w_i \in \mathcal{W}$, we have

$$\begin{aligned}\mathcal{OPT} = \sum_{j\in\mathcal{S}^*} c_j &= \sum_{w_i\in\mathcal{W}} \sum_{j\in\mathcal{W}_i} c_j + \sum_{j\in\mathcal{S}^*\cap\mathcal{W}} c_j \\ &\geq \sum_{w_i\in\mathcal{W}} \frac{c_{w_i}}{g} - O(\ln n),\end{aligned}$$

where the inequality holds because when n is large, $|\mathcal{W}| \ll n$.

Then the lemma holds. □

For LOG we have the following properties. The proofs are similar to those for LIN, and thus omitted.

Theorem 5. LOG *achieves computational efficiency, individual rationality, truthfulness, and $(\epsilon(e-1)/e, \delta)$-differential privacy, where $\epsilon > 0$ and $\delta \in (0, \frac{1}{2}]$ are two constants, e is the base of the natural logarithm. In addition, it has social cost at most $2^t H_m \mathcal{OPT}$ with probability at least $1 - e^{-t}$, for any constant*

$t > 0$ and $H_m = \sum_{j=1}^{m} 1/j$, where m is the number of sensing tasks, and $\mathcal{OPT}$ is the optimal social cost of the SCM problem.

Remarks: According to Theorem 4 in [30], the minimum weighted set cover problem can be reduced to the SCM problem. It is well known that the best-possible polynomial time approximation algorithm is an H_m-approximation algorithm for the weighted set cover problem [9], where H_m is the mth harmonic number. LOG has social cost at most $2^t H_m \mathcal{OPT}$, where t is a constant, and thus it is asymptotically optimal. Even though LIN cannot be proved to be asymptotically optimal in terms of the social cost, we will show in Section 6 that it achieves better privacy protection than LOG.

5 BidGuard-M: Differentially Private Auction Framework for the Multi-Bid Model

In this section, we design and analyze BidGuard-M, a differentially private auction framework for the multi-bid Model.

5.1 Design Rationale

BidGuard-M integrates the exponential mechanism with the reverse auction to achieve computational efficiency, individual rationality, truthfulness, differential privacy and approximate social cost minimization. In this framework, task-bid pairs are selected iteratively. In each iteration, one task is considered. Each of the task-bid pairs with this task is assigned a probability to be selected. The framework then selects one of them as the winning task-bid pair according to the probability distribution. Specifically, the probability of a task-bid pair to be selected is set according to a specific criterion. The above process repeats until all the sensing tasks can be completed by the selected task-bid pairs. Finally, the framework computes the payment to each winning task-bid pair.

5.2 Design of BidGuard-M

In this section, we will describe BidGuard-M in detail as illustrated in Algorithm 2.

BidGuard-M selects a winning task-bid pair for each task in $\mathcal{T}$ iteratively until all the tasks can be completed. All the winning task-bid pairs constitute $\mathcal{B}_W$. At the beginning of each iteration, BidGuard-M first selects for an unassigned task $t \in \mathcal{T}$ a set of task-bid pairs $\mathcal{B}_t$ in which $t_i^k = t$ for all $\beta_i^k \in \mathcal{B}_t$. BidGuard-M will assign each task-bid pair $\beta_i^k \in \mathcal{B}_t$ a probability to be selected as follows. It is desired to select the task-bid pair with the lowest b_i^k from $\mathcal{B}_t$. To apply the exponential mechanism, we need to design a score function, which is a non-increasing function of b_i^k. The probability of each task-bid pair to be selected is set according to the value of the score function. At last, BidGuard-M calculates the payment p_i^k for each winning task-bid pair $\beta_i^k \in \mathcal{B}_W$. Let $Pr_i(z)$ denote the probability of a task-bid pair being selected with bid z. According to Theorem 1, the payment to a winning task-bid pair is

$$p_i^k = b_i^k + \frac{\int_{b_i^k}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i^k)}.$$

For each user, if it has at least one winning task-bid pair, it is added into the winner set $\mathcal{S}$ and its payment $p_i = \sum_{\beta_i^k \in \mathcal{B}_i \cap \mathcal{B}_W} p_i^k$.

Algorithm 2. BidGuard-M

Input: A set of sensing tasks $\mathcal{T}$, a set of users $\mathcal{U}$, submitted task-bid profile $\overrightarrow{\mathcal{B}}$, and differential privacy parameter $\epsilon > 0$.
Output: A set of winners $\mathcal{S}$ and a payment profile $\overrightarrow{p}$.

$\mathcal{B}_W \leftarrow \emptyset$, $\mathcal{S} \leftarrow \emptyset$, $\mathcal{B}_t \leftarrow \emptyset$;
foreach $i \in \mathcal{U}$ **do** $p_i \leftarrow 0$
foreach $t \in \mathcal{T}$ **do**
  foreach $i \in \mathcal{U}$ **do**
    if $\exists \beta_i^k \in \mathcal{B}_i$ *such that* $t_i^k = t$ **then** $\mathcal{B}_t \leftarrow \mathcal{B}_t \cup \{\beta_i^k\}$
  end
  foreach $\beta_i^k \in \mathcal{B}_t$ **do**
    Calculate the probability $Pr_i(b_i^k)$ of each task-bid pair being selected according to the score function;
  end
  Select one task-bid pair randomly, denoted by $\beta_{i'}^{k'}$, according to the computed probability distribution;
  $\mathcal{B}_W \leftarrow \mathcal{B}_W \cup \{\beta_{i'}^{k'}\}$, $\mathcal{B}_t \leftarrow \emptyset$;
end
foreach $\beta_i^k \in \mathcal{B}_W$ **do**
  $p_i^k \leftarrow b_i^k + \frac{\int_{b_i^k}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i^k)}$;
end
foreach $i \in \mathcal{U}$ **do**
  if $\mathcal{B}_i \cap \mathcal{B}_W \neq \emptyset$ **then**
    $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$;
    $p_i \leftarrow \sum_{\beta_i^k \in \mathcal{B}_i \cap \mathcal{B}_W} p_i^k$;
  end
end
return $\mathcal{S}$ and $\overrightarrow{p}$.

5.3 Design of Score Functions

Same as the single-bid model, we adopt f_{LIN} and f_{LOG} as score functions. We will show that they have different theoretical bounds on the social cost (Section 5.4) and performance in simulations (Section 6).

Linear Score Function. For any task-bid pair $\beta_i^k \in \mathcal{B}_t$, the probability to be selected is

$$Pr_i(b_i^k) \propto \exp\left(\epsilon\left(1 - \frac{b_i^k}{b_{max}}\right)\right).$$

Note that in order to guarantee the value of the score function is nonnegative, we normalize b_i^k, i.e., $\frac{b_i^k}{b_{max}}$. Then the probability is

$$Pr_i(b_i^k) = \frac{\exp\left(\epsilon\left(1 - \frac{b_i^k}{b_{max}}\right)\right)}{\sum_{\beta_j^k \in \mathcal{B}_t} \exp\left(\epsilon\left(1 - \frac{b_j^k}{b_{max}}\right)\right)}. \quad (4)$$

Log Score Function. For any task-bid pair $\beta_i^k \in \mathcal{B}_t$, the probability to be selected is

$$Pr_i(b_i^k) \propto \exp\left(\epsilon \log_{1/2} \frac{b_i^k}{b_{max}}\right).$$

We also normalize the b_i^k, i.e., $\frac{b_i^k}{b_{max}}$ to guarantee the value of the score function is nonnegative. Then the probability is

$$Pr_i(b_i^k) = \frac{\exp\left(\epsilon \log_{1/2} \frac{b_i^k}{b_{max}}\right)}{\sum_{\beta_j^k \in \mathcal{B}_t} \exp\left(\epsilon \log_{1/2} \frac{b_j^k}{b_{max}}\right)}. \quad (5)$$

Throughout the rest of this paper, we denote the BidGuard-M with linear score function f_{LIN} and log score function f_{LOG} by LIN−M and LOG−M, respectively.

Illustrating Example. We use the example in Table 2 to illustrate how LIN−M works. Let $b_{min} = 1$, $b_{max} = 4$, and the differential privacy parameter $\epsilon = 0.1$. At the beginning, $\mathcal{T} = \{t_1, t_2, t_3\}$, $\mathcal{S} = \emptyset$, $\mathcal{B}_W = \emptyset$, $\mathcal{B}_t = \emptyset$, and $\mathcal{U} = \{1, 2, 3, 4, 5\}$. LIN−M starts to select users for every task in $\mathcal{T}$ iteratively. For t_1, LIN−M first constructs $\mathcal{B}_1 = \{\beta_1^1, \beta_2^1, \beta_3^1, \beta_4^1, \beta_5^1\}$. By (4), LIN−M calculates the probability of every task-bid pair in $\mathcal{B}_1$ to be selected. For example, the probability of β_1^1 to be selected is $Pr_1(1.5) = \exp(0.0625)/(\exp(0.0625) + \exp(0.075) + \exp(0.06) + \exp(0.025) + \exp(0.0375))$. LIN−M selects one task-bid pair based on the calculated probability distribution. Assume β_2^1 is selected, then $\mathcal{B}_W = \{\beta_2^1\}$. LIN−M executes the same process for t_2. We have $\mathcal{B}_2 = \{\beta_1^2, \beta_4^2\}$. The probabilities of β_1^2 and β_4^2 to be selected are $Pr_1(1.5) = \exp(0.0625)/(\exp(0.0625) + \exp(0.05))$ and $Pr_4(2) = \exp(0.05)/(\exp(0.0625) + \exp(0.05))$, respectively. Assume LIN−M selects β_1^2, then $\mathcal{B}_W = \{\beta_2^1, \beta_1^2\}$. For t_3, LIN−M constructs $\mathcal{B}_3 = \{\beta_3^2, \beta_5^2\}$. The probabilities of β_3^2 and β_5^2 to be selected are $Pr_3(2.4) = \exp(0.06)/(\exp(0.06) + \exp(0.0375))$ and $Pr_5(2.5) = \exp(0.0375)/(\exp(0.06) + \exp(0.0375))$, respectively. Assume LIN−M selects β_3^2, then $\mathcal{B}_W = \{\beta_2^1, \beta_1^2, \beta_3^2\}$. Once all tasks are assigned, LIN−M calculates the payment for each task-bid pair in $\mathcal{B}_W$. We have $p_2^1 = 1 + \frac{\int_1^4 Pr_2(z)dz}{Pr_2(1)}$, $p_1^2 = 1.5 + \frac{\int_{1.5}^4 Pr_1(z)dz}{Pr_1(1.5)}$, and $p_3^2 = 2.4 + \frac{\int_{2.4}^4 Pr_3(z)dz}{Pr_3(2.4)}$. At last, LIN−M calculates the winners set $\mathcal{S} = \{1, 2, 3\}$ and corresponding payments, i.e., $p_1 = p_1^2, p_2 = p_2^1$ and $p_3 = p_3^2$.

5.4 Analysis of BidGuard-M

In this section, we first analyze the properties of LIN−M.

Theorem 6. LIN−M *achieves computational efficiency, individual rationality, truthfulness, and* $2m\epsilon$*-differential privacy, where* $\epsilon > 0$ *is a constant and* m *is the number of sensing tasks. In addition, it has social cost at most* $\mathcal{OPT} + mO(\ln n)$ *with probability at least* $1 - 1/n^{O(1)}$*, where* $\mathcal{OPT}$ *is the optimal social cost of the SCM-M problem, and* n *is the number of users.*

Proof. We first prove the computational efficiency. The outer while-loop (Lines 4-13) will run at most m iterations since there are m tasks. Meanwhile, the two inner for-loops (Lines 4-6) and (Lines 7-9) will run at most n iterations since there are n users. The payment calculation for the winning task-bid pairs (Lines 13-15) will run at most m iterations since there are m tasks. The winner selection and payment calculation (Lines 16-21) will run at most n iterations since there are n users. Therefore, the total computational complexity of LIN−M is $O(mn)$. The individual rationality is guaranteed by the fact that the payment to each winning task-bid pair is $p_i^k = b_i^k + \frac{\int_{b_i^k}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i^k)} \geq b_i^k$. In order to prove the rest of this theorem, we prove the following lemmas. □

Lemma 4. LIN−M *is truthful.*

Proof. According to (4) and (5), the probability $Pr_i(b_i^k)$ of task-bid pair $\beta_i^k \in \mathcal{B}_t$ being selected in BidGuard-M is monotonically non-increasing in its bid b_i^k. In addition, no bid is greater than b_{max} in our model. Thus we have $\int_0^\infty Pr_i(z)dz = \int_0^{b_{max}} Pr_i(z)dz < \infty$. Furthermore, we have

$$\begin{aligned} &E[p_i^k] \\ &= \left(1 - Pr_i(b_i^k)\right) \times 0 + Pr_i(b_i^k) \times \left(b_i^k + \frac{\int_{b_i^k}^{b_{max}} Pr_i(z)dz}{Pr_i(b_i^k)}\right) \\ &= b_i^k Pr_i(b_i^k) + \int_{b_i^k}^{\infty} Pr_i(z)dz. \end{aligned}$$

Then, according to Theorem 1, the lemma holds. □

In order to quantify the differential privacy performance of LIN−M, we use the following lemmas.

Lemma 5 (Composability [33]). *The sequential application of randomized computation* M_i*, each giving* ϵ_i*-differential privacy, yields* $(\sum_i \epsilon_i)$*-differential privacy.*

Lemma 6. *For any constant* $\epsilon > 0$*,* LIN−M *achieves* $2m\epsilon$*-differential privacy, where* m *is the number of sensing tasks.*

Proof. Since LIN−M follows the exponential mechanism, it selects a task-bid pair for each task based on (4). According to Theorem 2, for each task LIN−M is 2ϵ-differential privacy, since the largest difference in the score function (Λ) is 1. LIN−M selects a task-bid pair for each task in $\mathcal{T}$ iteratively until all tasks can be finished. This is a sequential application of the selection mechanism for one task. Therefore, according to Lemma 5, LIN−M is $2m\epsilon$-differential privacy, since there are m tasks. □

Next, we bound the social cost of LIN−M.

Lemma 7. *With probability at least* $1 - 1/n^{O(1)}$*,* LIN−M *has social cost at most* $\mathcal{OPT} + mO(\ln n)$*, where* $\mathcal{OPT}$ *is the optimal social cost of the SCM-M problem,* m *is the number of sensing tasks and* n *is the number of users.*

Proof. Let $\mathcal{B}^*$ denote the optimal solution to the SCM-M problem. We denote as $\mathcal{B}_W$ an arbitrary set of winning task-bid pairs returned by LIN−M. Because only one task-bid pair is selected for each task and all tasks need to be completed, we have $|\mathcal{B}^*| = |\mathcal{B}_W|$. Therefore, for any task-bid pair $\beta_j^k \in \mathcal{B}^*$, there exists a task-bid pair $\beta_i^k \in \mathcal{B}_W$ such that $t_i^k = t_j^k$, and vice versa. According to Theorem 3, by taking $\alpha = O(\ln n)$, we have

$$b_i^k \leq b_j^k + O(\ln n), \quad (6)$$

with a probability of at least $1 - 1/n^{O(1)}$ for each task $t \in \mathcal{T}$. Summing (6) over all tasks, $\sum_{\beta_i^k \in \mathcal{B}_W} b_i^k \leq \sum_{\beta_j^k \in \mathcal{B}^*} b_j^k + mO(\ln n)$ with a probability at least $1 - 1/n^{O(1)}$. For truthful mechanisms, we have $b_i^k = c_i^k$ and $b_j^k = c_j^k$. Thus $\sum_{\beta_i^k \in \mathcal{B}_W} b_i^k$ is the social cost of LIN−M, and $\mathcal{OPT} = \sum_{\beta_j^k \in \mathcal{B}^*} b_j^k$.

This concludes the proof. □

For LOG−M we have the following properties. The proofs are similar to those for LIN−M, and thus omitted.

Theorem 7. LOG−M *achieves computational efficiency, individual rationality, truthfulness, and* $2m \log_{\frac{1}{2}}(\frac{1}{1+\Delta})\epsilon$*-differential*

privacy, where m is the number of sensing tasks, Δ is the maximum difference in the bidding price, and $\epsilon > 0$ is a constant. In addition, it has social cost at most $2^t\mathcal{OPT}$ with probability at least $1 - e^{-t}$, for any constant $t > 0$ and $\mathcal{OPT}$ is the optimal social cost of the SCM-M problem.

Remarks: LOG−M has social cost at most $2^t\mathcal{OPT}$, where t is a constant, and thus it is asymptotically optimal.

6 Performance Evaluation

In this section, we evaluate the performance of BidGuard and BidGuard-M and compare them respectively with TRAC [15] and DP-hSRC [23]. TRAC is closest to our work in terms of the design objective, but does not protect users' bid privacy. DP-hSRC considers users' bid privacy, but minimizes total payment instead of social cost.

6.1 Simulation Setup

All the evaluation results are based on a real data set of taxi traces. The dataset consists of the traces of 320 taxi drivers, who work in the center of Rome [6]. Each taxi driver has a tablet that periodically (every 7s) retrieves the GPS locations (latitude and longitude) and sends it with the corresponding driver ID to a central server. The mobility pattern of taxi traces can be used to depict the mobility of smartphone users as in [25], [47].

We consider a mobile crowdsensing system where the task is to measure the cellular signal strength at specific locations. Each user can sense the cellular signal strength within the area centered at the user's location with a radius of 30 m. Tasks are represented by GPS locations reported by taxis. We assume that the driver of each taxi is a user. We preprocess the tasks such that each task can be sensed by at least two users according to our system model.

We use three metrics to evaluate the performance: *social cost*, *total payment* and *privacy leakage*. The social cost, as defined in Section 3, refers to the total cost of all selected users. The total payment measures the payment paid by the platform to all selected users. We first compare the social cost and total payment of BidGuard and BidGuard-M with TRAC. Then we compare the social cost of BidGuard with the optimal social cost. We define *privacy leakage* to quantitatively measure the differential privacy performance of BidGuard and BidGuard-M.

Privacy Leakage. Given a mechanism M, let $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$ be two task-bid profiles, which only differ in one user's bid. Let $M(\overrightarrow{\beta})$ and $M(\overrightarrow{\beta'})$ denote the outcome of M with input $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$, respectively. The privacy leakage, denoted by PL, is defined as the Kullback-Leibler divergence of the two outcome probability distributions based on $\overrightarrow{\beta}$ and $\overrightarrow{\beta'}$,

$$PL = \sum_{o \in \mathcal{O}} Pr\left[M(\overrightarrow{\beta}) = o\right] \ln\left(\frac{Pr\left[M(\overrightarrow{\beta}) = o\right]}{Pr\left[M(\overrightarrow{\beta'}) = o\right]}\right). \quad (7)$$

Note that the smaller the PL value is, the harder it is to distinguish the two task-bid profiles, and thus the better the privacy preserving performance is achieved.

In our evaluation, we randomly select locations as the sensing tasks according to the settings. We assume the bids of users are randomly distributed over $[1, 50]$ for BidGuard and $[1, 10]$ for BidGuard-M. Because users in BidGuard bid for a set of tasks, while users in BidGuard-M bid for a single task. We generate users' bids according to two different distributions, i.e., uniform distribution and normal distribution. To evaluate the impact of the number of sensing tasks on the performance metrics, we set the number of users to 200 and vary the number of sensing tasks from 20 to 60 with a step of 10. To evaluate the impact of the number of users on the performance metrics, we set the number of sensing tasks to 150 and vary the number of users from 100 to 300 with a step of 50. For the differential privacy parameters, we set $\epsilon = 0.1$ and $\delta = 0.25$ as default. All the results are averaged over 1,000 independent runs for each setting. Note that since the performances under both uniform and normal distributions follow the same pattern according to our evaluation, in the following we only show the performance under the uniform distribution.

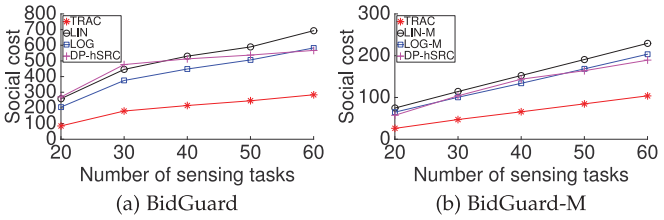


Fig. 1. Impact of the number of sensing tasks on the social cost.

6.2 Evaluation of Social Cost

We first compare the social cost of BidGuard and BidGuard-M with that of TRAC and DP-hSRC. Note that TRAC is optimal in the multi-bid model. The impact of the number of sensing tasks on the social cost of BidGuard and that of BidGuard-M is shown in Figs. 1a and 1b, respectively. We observe that the social cost of TRAC, DP-hSRC, BidGuard and BidGuard-M all increase when the number of sensing tasks grows. This is because with more sensing tasks, the platform may select more users incurring a higher social cost. We also see that the social cost of TRAC is lower than those of DP-hSRC, BidGuard and BidGuard-M. This is because, in each iteration, TRAC is determinate to select the user with the lowest criterion value (defined in (1)) in the single-bid model and the user with the lowest bid for each task in the multi-bid model. In contrast, since both BidGuard and BidGuard-M are randomized, they cannot always guarantee to select the user with the lowest criterion value or the lowest bid in each iteration. DP-hSRC selects users based on a threshold price and has no performance guarantee on the social cost. Besides, the social cost of LOG is smaller than that of LIN, and the social cost of LOG−M is lower than that of LIN−M. This is because, both LOG and LOG−M prefer to select users with low bid, as the log score function will give more probability of being selected to low-bid users.

Figs. 2a and 2b depict the impact of the number of users on the social cost of BidGuard and BidGuard-M, respectively. We see that the social cost decreases slightly when the number of users increases for TRAC, DP-hSRC, BidGuard and BidGuard-M. This is because, with more users, the platform can find more low-cost users to complete the sensing tasks. The social cost of TRAC is lower than those of DP-hSRC, BidGuard and BidGuard-M. The reason is same

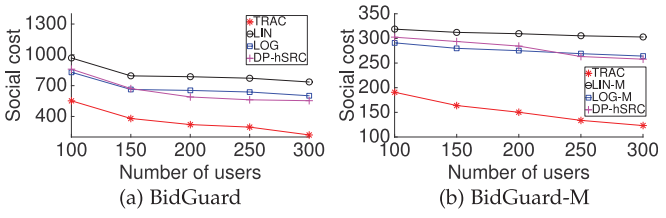


Fig. 2. Impact of the number of users on the social cost.

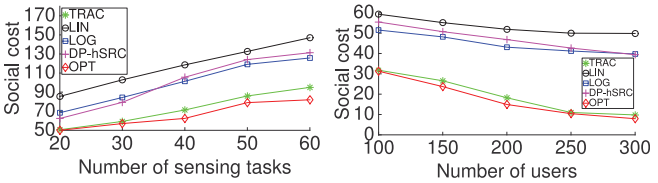


Fig. 3. Comparison of BidGuard, TRAC, DP-hSRC, and OPT.

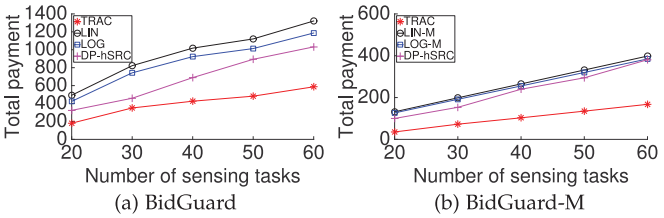


Fig. 4. Impact of the number of sensing tasks on the total payment.

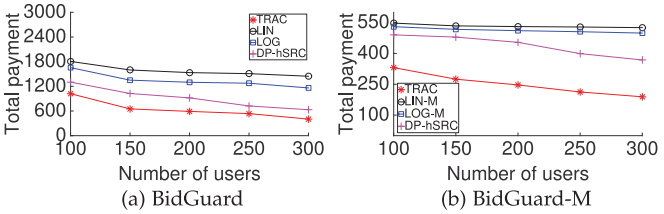


Fig. 5. Impact of the number of users on the total payment.

as explained for Fig. 1. Meanwhile, for the same reason as above, the social cost of LOG is lower than that of LIN, and the social cost of LOG−M is lower than that of LIN−M.

In Fig. 3, we compare the social cost of incentive mechanisms in the single-bid model. Let OPT denote the optimal solution. Since finding the optimal solution takes exponential time, we set the number of the users to 10 for Fig. 3a, and set the number of sensing tasks to 4 for Fig. 3b. We observe that Figs. 3a and 3b have the same pattern in Figs. 1a and 2a, respectively. The reason is similar to those explained for Figs. 1a and 2a. Furthermore, we observe that BidGuard sacrifices the social cost for the users' bid privacy, compared to TRAC and the optimal solution. Note that in Fig. 3b, the social cost of TRAC is very close to that of OPT. This is because TRAC is an H_K-approximation algorithm, where $H_k \approx 2.34$ in this figure.

6.3 Evaluation of Total Payment

In Figs. 4 and 5, we plot the impact of the number of sensing tasks and the impact of the number of users on the total payment of BidGuard and BidGuard-M, respectively. The results show that the total payment of TRAC, DP-hSRC, BidGuard and BidGuard-M all follow the same pattern as the social cost. In addition, both LOG and LOG−M have lower total payment than LIN and LIN−M, respectively. This is because the log score function could select users with lower bids as shown in Figs. 1 and 2. We also observe that the total payment of DP-hSRC is lower than BidGuard and BidGuard-M. This is because DP-hSRC selects and pays users according to a single-price, and thus it has performance guarantee on the total payment. However, DP-hSRC can only achieve approximate truthfulness, which ensures that no user is able to make more than a slight gain in its expected utility by bidding untruthfully. In addition, in the next section, we will see that the privacy protection of LIN and LIN−M are better than that of DP-hSRC.

6.4 Evaluation of Privacy Leakage

Next, we evaluate BidGuard and BidGuard-M in terms of privacy leakage. Figs. 6a and 7a plot the impact of the number of sensing tasks on the privacy leakage for BidGuard and BidGuard-M, respectively. Figs. 6b and 7b plot the impact of the number of users on the privacy leakage for BidGuard and BidGuard-M, respectively.

We observe that the privacy leakage values of BidGuard, BidGuard-M and DP-hSRC are very small. This is because they all achieve differential privacy. However, the privacy leakage value of TRAC is positive infinity, which indicates a bad differential privacy performance. This is because TRAC does not protect users' bid privacy, and the denominator could be 0 for TRAC according to (7).

In both Figs. 6a and 7a, we see that the privacy leakage of both LIN and LIN−M are always smaller than that of LOG and LOG−M, respectively, which indicates that LIN and LIN−M have better privacy protection performance than LOG and LOG−M, respectively. This is because the linear score function treats the probability of every outcome uniformly, however, the log score function gives more probability to the outcome with low social cost. We also observe that the privacy leakage of both LIN and LIN−M are always

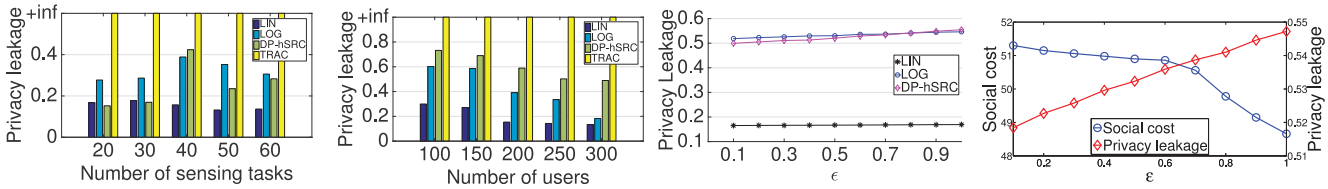
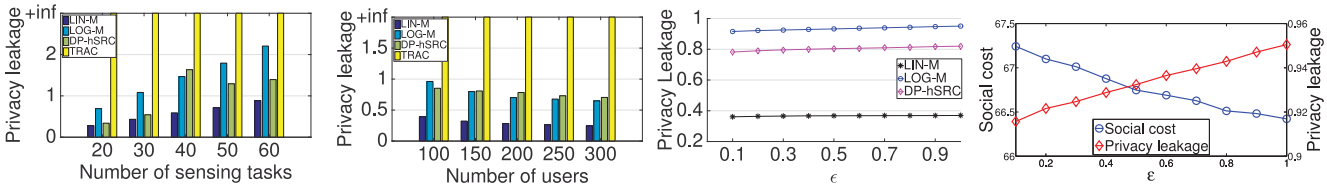


Fig. 6. Evaluation of privacy leakage for BidGuard.



(a) Impact of the number of sensing tasks (b) Impact of the number of users (c) Impact of ϵ (d) Social cost v.s. privacy leakage

Fig. 7. Evaluation of privacy leakage for BidGuard-M.

smaller than that of DP-hSRC. This is because DP-hSRC is a single-price mechanism in which one user's bid change may significantly change the outcome of the mechanism, and thus increases *PL* according to (7). We do not observe a pattern of the privacy leakage when the number of tasks increases for BidGuard. The reason is, according to the definition of privacy leakage, the difference between the probabilities of two outcomes to be selected is independent of the number of sensing tasks. However, the privacy leakage value increases when the number of tasks increases for BidGuard-M. This is because, according to Theorems 6 and 7, the differential privacy performance of BidGuard-M is reversely linear to the number of sensing tasks.

In both Figs. 6b and 7b, we can see the impact of the number of users on the privacy leakage of BidGuard and BidGuard-M, respectively. Note that the privacy leakage value decreases when the number of users increases for both DP-hSRC, BidGuard and BidGuard-M. This is because the probability of each outcome decreases as the number of users increases. Specifically, the more users in the system, the more possible outcomes for DP-hSRC, BidGuard and BidGuard-M, the less difference between the probabilities of two outcomes to be selected, and thus the better differential privacy performance. We also see the privacy protection performance of LIN and LIN−M are better than that of LOG and LOG−M, respectively. In addition, the privacy protection performance of LIN and LIN−M are also better than that of DP-hSRC. The reason for this is similar to that discussed before.

Figs. 6c and 7c show the impact of the differential privacy parameter ϵ on the privacy leakage for BidGuard and BidGuard-M, respectively. The results show that the value of ϵ has more impact on the privacy leakage for LOG and LOG−M than that of LIN and LIN−M, respectively. This is because the log score function is more sensitive than the linear score function. For LOG and LOG−M, the privacy leakage increases slightly when the value of ϵ grows. This is because, theoretically, the larger the ϵ is, the worse the differential privacy is achieved, and thus the higher privacy leakage. Meanwhile, it is easy to observe that the privacy leakage of LIN and LIN−M are smaller than that of DP-hSRC, LOG and LOG−M, respectively. This can also be explained by the same reason for Fig. 6a.

Figs. 6d and 7d illustrate the tradeoff between the social cost and the privacy leakage of LOG and LOG−M, respectively. We observe that the privacy leakage decreases as the decreasing of ϵ. The reason is similar to that discussed for Fig. 6c. However, this improvement in privacy comes at a cost of the increased social cost for both LOG and LOG−M.

Remarks: Compared with TRAC, which does not protect users' bid privacy, both BidGuard and BidGuard-M sacrifice the social cost and payment for the users' bid privacy. Compared with DP-hSRC, BidGuard and BidGuard-M have better bid privacy preservation and lower social cost in most cases although incurring higher total payment. In addition, BidGuard and BidGuard-M achieve truthfulness while DP-hSRC achieves approximate truthfulness. Besides, LIN and LIN−M outperform LOG and LOG−M in terms of privacy protection, respectively. However LOG and LOG−M have lower social cost and payment.

7 Conclusion and Future Work

In this paper, we have proposed two general frameworks, BidGuard and BidGuard-M, for privacy-preserving mobile crowdsensing incentive mechanisms, which achieve computational efficiency, individual rationality, truthfulness, approximate social cost minimization, and differential privacy. We designed two score functions, linear and log, to realize the frameworks. Note that, both BidGuard and BidGuard-M with log function are asymptotically optimal in terms of the social cost. We evaluated the performance of our frameworks through extensive simulations.

In the future, we plan to design different score functions which might have better performance in terms of differential privacy or proximate social cost minimization. In addition, we plan to evaluate our frameworks by using real-world experiments.

Acknowledgments

This is an extended and enhanced version of the paper [30] that appeared in IEEE CNS 2016. This research was supported in part by US National Science Foundation grants 1444059, 1461886, 1717315, and 1717197, NSFC grant 61472193, NSF of Jiangsu Province BK20141429, and CCF-Tencent RAGR20150107.

References

[1] Auction.com. (2017). [Online]. Available: https://www.auction.com/
[2] eBay. (2017). [Online]. Available: https://www.ebay.com/
[3] FCC auctions releases. (2017). [Online]. Available: http://wireless.fcc.gov/auctions/default.htm?job=archived_releases
[4] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2013, pp. 901–914.
[5] A. Archer and É. Tardos, "Truthful mechanisms for one-parameter agents," in *Proc. IEEE Symp. Found. Comput. Sci.*, 2001, pp. 482–491.
[6] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, and A. Rabuffi, "CRAWDAD data set roma/taxi (v. 2014-07-17)," Jul. 2014. [Online]. Available: http://crawdad.org/roma/taxi/
[7] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan, "Truthful mechanisms for agents that value privacy," in *Proc. ACM Conf. Electron. Commerce*, 2013, pp. 215–232.

[8] D. Christin, "Privacy in mobile participatory sensing: Current trends and future challenges," *J. Syst. Softw.*, vol. 116, pp. 57–68, 2016.
[9] V. Chvatal, "A greedy heuristic for the set-covering problem," *Math. Operations Res.*, vol. 4, no. 3, pp. 233–235, 1979.
[10] C. Clifton, "Using sample size to limit exposure to data mining," *J. Comput. Secur.*, vol. 8, no. 4, pp. 281–307, 2000.
[11] E. De Cristofaro and C. Soriente, "Short paper: PEPSI—privacy-enhanced participatory sensing infrastructure," in *Proc. ACM Conf. Wireless Netw. Secur.*, 2011, pp. 23–28.
[12] C. Dwork, "Differential privacy," in *Encyclopedia of Cryptography and Security*. Berlin, Germany: Springer, 2011, pp. 338–340.
[13] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. 3rd Conf. Theory Cryptography*, 2006, pp. 265–284.
[14] J. Fan, Q. Li, and G. Cao, "Privacy-aware and trustworthy data aggregation in mobile sensing," in *Proc. IEEE Conf. Commun. Netw. Secur.*, 2015, pp. 31–39.
[15] 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 *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 1231–1239.
[16] H. Gao, et al., "A survey of incentive mechanisms for participatory sensing," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 918–943, Apr.–Jun. 2015.
[17] B. Gedik and L. Liu, "Protecting location privacy with personalized k-anonymity: Architecture and algorithms," *IEEE Trans. Mobile Comput.*, vol. 7, no. 1, pp. 1–18, Jan. 2008.
[18] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy, and incentive provision for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 839–853, Oct. 2016.
[19] A. Gupta, K. Ligett, F. McSherry, A. Roth, and K. Talwar, "Differentially private combinatorial optimization," in *Proc. Annu. ACM-SIAM Symp. Discrete Algorithms*, 2010, pp. 1106–1125.
[20] T. H. Hinke, H. S. Delugach, and R. P. Wolf, "Protecting databases from inference attacks," *Comput. Secur.*, vol. 16, no. 8, pp. 687–708, 1997.
[21] Z. Huang and S. Kannan, "The exponential mechanism for social welfare: Private, truthful, and nearly optimal," in *Proc. IEEE Symp. Found. Comput. Sci.*, 2012, pp. 140–149.
[22] S. Jajodia and C. Meadows, "Inference problems in multilevel secure database management systems," *Inf. Secur.: An Integrated Collection Essays*, vol. 1, pp. 570–584, 1995.
[23] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2016, pp. 344–353.
[24] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "INCEPTION: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proc. ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2016, pp. 341–350.
[25] X. Jin and Y. Zhang, "Privacy-preserving crowdsourced spectrum sensing," in *Proc. IEEE Conf. Comput. Commun.*, 2016, pp. 1–9.
[26] P. Kalnis, G. Ghinita, K. Mouratidis, and D. Papadias, "Preventing location-based identity inference in anonymous spatial queries," *IEEE Trans. Knowl. Data Eng.*, vol. 19, no. 12, pp. 1719–1733, Dec. 2007.
[27] I. Krontiris and T. Dimitriou, "A platform for privacy protection of data requesters and data providers in mobile sensing," *Comput. Commun.*, vol. 65, pp. 43–54, 2015.
[28] Q. Li and G. Cao, "Providing privacy-aware incentives for mobile sensing," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun.*, 2013, pp. 76–84.
[29] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proc. IEEE Int. Conf. Distrib. Comput. Syst.*, 2014, pp. 208–217.
[30] J. Lin, D. Yang, M. Li, J. Xu, and G. Xue, "BidGuard: A framework for privacy-preserving crowdsensing incentive mechanisms," in *Proc. IEEE Conf. Commun. Netw. Secur.*, 2016, pp. 439–447.
[31] X. Liu, Z. Zhou, W. Diao, Z. Li, and K. Zhang, "When good becomes evil: Keystroke inference with smartwatch," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2015, pp. 1273–1285.
[32] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *Proc. IEEE Conf. Comput. Commun.*, 2014, pp. 127–135.
[33] F. McSherry and K. Talwar, "Mechanism design via differential privacy," in *Proc. IEEE Symp. Found. Comput. Sci.*, 2007, pp. 94–103.
[34] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *Proc. ACM Conf. Embedded Netw. Sens. Syst.*, 2008, pp. 323–336.
[35] X. Niu, M. Li, Q. Chen, Q. Cao, and H. Wang, "EPPI: An E-cent-based privacy-preserving incentive mechanism for participatory sensing systems," in *Proc. IEEE 33rd Int. Perform. Comput. Commun. Conf.*, 2014, pp. 1–8.
[36] H. Ohashi, "Effects of transparency in procurement practices on government expenditure: A case study of municipal public works," *Rev. Ind. Org.*, vol. 34, no. 3, pp. 267–285, 2009.
[37] S. Reporter, "The proof is in the profit. Auction transparency works." (2017). [Online]. Available: http://dealerbidsale.com/the-proof-is-in-the-profit-auction-transparency-works/
[38] F. Restuccia, S. K. Das, and J. Payton, "Incentive mechanisms for participatory sensing: Survey and research challenges," *ACM Trans. Sens. Netw.*, vol. 12, no. 2, 2016, Art. no. 13.
[39] M. Shin, C. Cornelius, D. Peebles, A. Kapadia, D. Kotz, and N. Triandopoulos, "AnonySense: A system for anonymous opportunistic sensing," *Pervasive Mobile Comput.*, vol. 7, no. 1, pp. 16–30, 2011.
[40] A. Stoica and C. Farkas, "Secure XML views," *Res. Directions Data Appl. Secur.*, vol. 128, pp. 133–146, 2003.
[41] J. Sun and H. Ma, "Privacy-preserving verifiable incentive mechanism for online crowdsourcing markets," in *Proc. Int. Conf. Comput. Commun. Netw.*, 2014, pp. 1–8.
[42] H. To, G. Ghinita, and C. Shahabi, "A framework for protecting worker location privacy in spatial crowdsourcing," *Proc. VLDB Endowment*, vol. 7, no. 10, pp. 919–930, 2014.
[43] Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong, and G. Wu, "An incentive mechanism with privacy protection in mobile crowdsourcing systems," *Comput. Netw.*, vol. 102, pp. 157–171, 2016.
[44] T. Wen, Y. Zhu, and T. Liu, "P2: A location privacy-preserving auction mechanism for mobile crowd sensing," in *Proc. IEEE Global Commun. Conf.*, 2016, pp. 1–6.
[45] D. Xiao, "Is privacy compatible with truthfulness?" in *Proc. Conf. Innovations Theoretical Comput. Sci.*, 2013, pp. 67–86.
[46] J. Xu, J. Xiang, and Y. Li, "Incentivize maximum continuous time interval coverage under budget constraint in mobile crowd sensing," *Wireless Netw.*, vol. 23, no. 5, pp. 1549–1562, 2017.
[47] J. Xu, J. Xiang, and D. Yang, "Incentive mechanisms for time window dependent tasks in mobile crowdsensing," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6353–6364, Nov. 2015.
[48] D. Yang, G. Xue, G. Fang, and J. Tang, "Incentive mechanisms for crowdsensing: Crowdsourcing with smartphones," *IEEE/ACM Trans. Netw.*, vol. 24, no. 3, pp. 1732–1744, Jun. 2016.
[49] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. Annu. Int. Conf. Mobile Comput. Netw.*, 2012, pp. 173–184.
[50] B. Zhang, et al., "Privacy-preserving QoI-aware participant coordination for mobile crowdsourcing," *Comput. Netw.*, vol. 101, pp. 29–41, 2016.
[51] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, "Truthful incentive mechanisms for crowdsourcing," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 2830–2838.
[52] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3190–3200, Dec. 2014.
[53] P. Zhou, Y. Zheng, and M. Li, "How long to wait?: Predicting bus arrival time with mobile phone based participatory sensing," in *Proc. Int. Conf. Mobile Syst. Appl. Services*, 2012, pp. 379–392.
[54] R. Zhu and K. G. Shin, "Differentially private and strategy-proof spectrum auction with approximate revenue maximization," in *Proc. IEEE Conf. Comput. Commun.*, 2015, pp. 918–926.

Jian Lin (S'15) received the BS and MS degree in computer science from Southwest Jiaotong University, Chengdu, China, in 2011 and 2014, respectively. He has been working toward the PhD degree in computer science with Colorado School of Mines, Golden, Colorado, since 2014. His research interests include economic and optimization approaches to networks, mobile crowdsensing, and mobile computing. He is a student member of the IEEE.

Dejun Yang (M'13) received the BS degree from Peking University, Beijing, China, in 2007 and the PhD degree in computer science from Arizona State University, Tempe, Arizona, in 2013. Currently, he is the Ben L. Fryrear assistant professor of computer science with the Colorado School of Mines, Golden, Colorado. His research interests include economic and optimization approaches to networks, crowdsourcing, smart grid, and security and privacy. He has served as a technical program committee member for many conferences, including the IEEE International Conference on Computer Communications (INFOCOM), the IEEE International Conference on Communications (ICC), and the IEEE Global Communications Conference (GLOBECOM). He has received Best Paper Awards at IEEE GLOBECOM (2015), the IEEE International Conference on Mobile Ad hoc and Sensor Systems (2011), and IEEE ICC (2011 and 2012), as well as a Best Paper Award Runner-up at the IEEE International Conference on Network Protocols (2010). He is a member of the IEEE.

Ming Li (S'15) received the BS degree in geochemistry from Peking University, Beijing, China, in 2009 and the MS degree in computer science from the Colorado School of Mines, Golden, Colorado, in 2015. She is currently working toward the PhD degree at the Colorado School of Mines. Her main research interests include game theory, contract theory, crowdsourcing, smartphone-based sensing systems, and visible light communication technologies. She is a student member of the IEEE.

Jia Xu (M'15) received the MS degree from the School of Information and Engineering, Yangzhou University, Jiangsu, China, in 2006 and the PhD degree from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Jiangsu, China, in 2010. He is currently a professor in the Nanjing University of Posts and Telecommunications. He was a visiting scholar in the Department of Electrical Engineering & Computer Science, Colorado School of Mines from Nov. 2014 to May 2015. His main research interests include crowdsourcing, opportunistic networks, and wireless sensor networks. He is a member of the IEEE.

Guoliang Xue (M'96-SM'99-F'11) received the BS degree in mathematics and the MS degree in operations research from Qufu Normal University, Qufu, China, in 1981 and 1984, respectively, and the PhD degree in computer science from the University of Minnesota, Minneapolis, Minnesota, in 1991. Currently, he is a professor of computer science with Arizona State University, Tempe, Arizona. His research interests span the areas of Quality of Service provisioning, network security and privacy, crowdsourcing and network economics, RFID systems and Internet of Things, smart city, and smart grids. He is the elected vice president of the IEEE Communications Society for Conferences (2016-2017). He served as a technical program committee (TPC) co-chair for the IEEE International Conference on Computer Communications (INFOCOM) (2010) and as an editor for the *IEEE/ACM Transactions on Networking* (2010-2014). He regularly serves on the technical program committee of security conferences such as the ACM Symposium on Information, Computer, and Communications Security; the ACM Conference on Computer and Communications Security; and the IEEE Conference on Communications and Network Security, as well as networking conferences such as the IEEE International Conference on Network Protocols, IEEE INFOCOM, and the ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc). He is the area editor of the *IEEE Transactions on Wireless Communications*, overseeing the wireless networking area. He is a fellow of the IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.