If You Do Not Care About It, Sell It: Trading Location Privacy in Mobile Crowd Sensing

Wenqiang Jin*, Mingyan Xiao*, Ming Li*, Linke Guo†,
* The University of Texas at Arlington, U.S.
† The State University of New York at Binghamton, U.S.
Email: wenqiang.jin@mavs.uta.edu, mingyan.xiao@mavs.uta.edu, ming.li@uta.edu, lguo@binghamton.edu

***Abstract*—Mobile crowd sensing (MCS) is a technique where sensing tasks are outsourced to a crowd of mobile users. Since most of sensing tasks are location-dependent, workers are required to embed their locations into sensing reports, which incurs location privacy vulnerabilities. Realizing that workers perceive their location privacy differently, in this work we construct an auction-based trading market, facilitating location privacy trading between workers and the platform. Each worker can decide how much location privacy to disclose to the platform based on its own location privacy leakage budget ξ. The higher ξ is, the less secrecy its reported location preserves. As a result, it receives higher payment from the platform as a compensation to its privacy loss. Besides, our mechanism enables the platform to select a suitable set of winning workers to achieve desirable service accuracy. For this purpose, a heuristic algorithm is devised, with polynomial-time complexity and bounded optimality gap. As formally proved in this manuscript, our proposed mechanism guarantees a series of nice properties, including *ξ-privacy*, (α, β)-*accuracy*, and *budget feasibility*.**

***Index Terms*—Location privacy; mobile crowd sensing; privacy trading**

I. Introduction

A. Motivation

Mobile crowd sensing (MCS) emerges as a promising sensing paradigm that outsources the collection of data to a crowd of participating users, namely workers, with mobile devices, which are equipped with a plethora of on-board sensors (e.g., compass, accelerometer, gyroscope, camera, GPS) to capture data from surrounding environment. A large variety of MCS systems have been deployed, including noise mapping, smart transportation, road surface monitoring, indoor floor plan reconstruction, healthcare, and many others. Sensing data in these MCS applications are mostly location dependent. For example, in noise mapping, the distribution of the urban noise varies according to different geographic areas; for the case of realtime traffic maps, the traffic volume is tied to a specific section of road. Therefore, workers are typically required to embed sensing locations, i.e., their coordinates, in each piece of sensing report, which may, however, cause privacy breach. Thus, protecting location privacy is essential to attract participating workers in MCS. While there has been some existing research on this topic [1]–[5], most of them adopt conventional approaches, such as cloaking and k-anonymity.

On the other hand, individuals may perceive their privacy differently; Some may impose stringent requirement over privacy leakage, while some others accept monetary reward in trade of their personal data. From the coupons offered for revealing opinions of a product to the large-scale trade of personal information by data brokers such as Acxiom [6], the commoditization of private data has been trending up. Some theoretical research has been devoted on private data trading. However, most of them, including [7]–[11], focus on scenarios where there exists an "agent" for individual users to trade their privacy with data buyers. In another word, the "agent" is assumed trustworthy and users do not have control over their own data. In this concern, Wang et al. [12], [13] proposed to take over privacy control from "agents" and return it to individual users. In their design, users make independent decisions on how much privacy to disclose to data buyers through adding different amount of noise to their original data. Clearly, this approach would inevitably impact the accuracy of the data aggregation results. Yet, they fail to provide a precise measurement on such an accuracy degradation. Besides, data buyers cannot freely choose what accuracy level of data to purchase. All these limitations render their contributions of less practical use.

B. Design Rationality

Inspired by [12], [13], in this paper we develop a trading market for MCS. Not only does it help the platform to recruit workers, as what traditional MCS markets do, but also facilitates workers to sell location privacy. More importantly, the platform can decide which subset of workers to pick, depending on how noisy their reported locations are, so as to provide accuracy guaranteed MCS services. Nonetheless, this is not an easy task.

First and foremost, in order to facilitate the trading, it is important to quantify location privacy. In this regard, the commoditization of location privacy has dovetailed nicely with the development of the theoretical underpinnings of it: recent work on *geo-indistinguishability* [14] provides a compelling definition and a precise way to quantify its sale. Owing to different perception on location privacy, each worker has its own *location privacy leakage budget* ξ. The larger ξ is, the loose requirement a worker has. In another word, this worker is more willing to sell its location privacy.

In order to stimulate workers to participate in task sensing, the platform has to compensate them for their sensing cost and privacy loss. However, in practice, the platform is usually limited with its monetary budget. Such a constraint can significantly impact trading outcomes. For example, even there are sufficient workers willing to sell their location privacy, the platform may be short of budget to afford such cost. Therefore, *budget feasibility* needs be taken into account in the mechanism design.

To enable the platform to quantify service accuracy so as to pick suitable worker set, a novel location obfuscation mechanism is devised. Each worker chooses its obfuscated location to report in a probabilistic manner. We further define (α, β)-*accuracy*, $\Pr[loss \leq \alpha] \geq \beta$, where $loss$ is the service degradation caused by workers' location manipulation and β is a confidence level. The introduction of (α, β)-*accuracy* bridges worker location privacy and MCS service accuracy.

C. Related Work

Protecting location privacy in MCS has attracted increasing attention. According to a recent survey on the MCS privacy issues [15], cloaking is one of the most widely used strategies in practice, e.g., [1]–[3]. Besides, k-anonymity [16] and location obfuscation [4], [5] have also been investigated in this domain. These works treat location privacy from different workers equally. In fact, individuals may perceive different values towards their privacy. Thus, in this work we endeavor to provide workers with more flexibility in selling their location information and determining their privacy level.

Data privacy has also been studied in the context of MCS, e.g., [17]–[19]. Their main idea is to protect worker's reported data from the platform during data aggregation, since individual data can potentially disclose sensitive information regarding their reporters. Techniques, such as cryptographic multi-party computation, data perturbation and anonymization have been employed. Recently, Jin et al. [20] incorporated workers' privacy cost into the incentive mechanism. Note that its objective is to prevent outsiders from identifying individual data, where the platform is assumed trustworthy. Instead, like [21], in this work we aim to protect worker's privacy from the platform. While [21] allows workers to add noise to their original data and get paid accordingly, users cannot customize how much noise to add; the noise distribution is determined by the platform.

Treating user privacy as commodities, Ghosh and Roth [7] are among the first to lay a theoretical foundation for selling private data. [8]–[11] also fall into this line of research. However, these works assume the existence of a trustworthy "agent" to sell user's data. Very recent works [12], [13] propose private data trading that users take full control of their own data. Nonetheless, they adopt game-theoretic models, which, however, may end up with an inefficient equilibrium; the accuracy of data aggregation is not guaranteed.

The rest of this paper is organized as follows. In Section II, we give a system overview. Design objectives of this work are described in Section III. Details of our mechanism design are discussed in Section IV. The analysis over the properties achieved in MCS markets are conducted in Section V. Extensive simulation results are provided in Section VI. Finally, we conclude the paper in Section VII.

II. System Overview

A. MCS Systems

We consider a general MCS system consisting of a platform and set of participating workers, denoted as $\mathcal{W} = \{w_1, \cdots, w_j, \cdots, w_M\}$. The platform publishes a set of sensing tasks $\mathcal{T} = \{\tau_1, \cdots, \tau_i, \cdots, \tau_N\}$. Since sensing tasks are generally location dependent, workers are required to indicate their sensing location in each sensing report. However, it allows the platform to track workers and is thus privacy compromising. To protect worker's location privacy from the platform, in our mechanism a worker w_j is allowed to report an obfuscated location z_j other than its genuine one l_j. Since workers value their location privacy differently, each of them independently chooses its own privacy budget ξ_j, which indicates the maximum privacy w_j is willing to disclose. Generally, the lower ξ_j is, the more stringent requirement w_j imposes over its location privacy. As a result, less useful location information z_j leaks to the platform that assists the inference of l_j. Since obfuscated locations affect the accuracy of data aggregation, the platform needs to decide which subset of workers to select for task sensing so as to provide accuracy-guaranteed MCS services. The framework of our auction based MCS market is summarized as follows.

- The platform announces a set of sensing tasks $\mathcal{T}$ to workers.
- Then, each worker $w_j \in \mathcal{W}$ submits its interested sensing tasks $\mathcal{T}_j$ and bid b_j to the platform, where the bid reflects w_j's valuation of privacy loss and sensing costs (see Section II-D). Following certain criterion (see Section IV-A), the platform determines a winner set $\mathcal{W}^*$, i.e., the workers to fulfill sensing tasks, and their payments $\boldsymbol{p}$.
- Each winning worker $w_j \in \mathcal{W}^*$ then conducts sensing tasks $\mathcal{T}_j$, prepares its sensing reports with obfuscated location z_j, and forwards them to the platform.
- Finally, the platform aggregates over collected sensing reports, derives sensing results, and publishes them to the community or sends back to task requestors.

B. Geo-information Quality Model

To hide the exact location l_j from the platform, each worker $w_j \in \mathcal{W}$ adopts a location obfuscation mechanism (see Section IV-B) to convert l_j into an obfuscated location z_j. Then, instead of l_j, worker w_j reports z_j together with its sensing data. As a result, its sensing report will inevitably experience accuracy degradation, so does the final aggregation result derived at the platform. Because this information loss is caused by worker's location manipulation, we refer it as *geo-information loss*, which is defined by

$$loss = \sum_{j: w_j \in \mathcal{W}^*} d(l_j, z_j) \tag{1}$$

where $d(l_j, z_j)$ represents w_j's drift distance. *Geo-information loss* is calculated as a sum over all winning workers' drift distances. A larger value of $loss$ implies that more "noises" are added to workers' reported locations, which results in poorer service accuracy MCS provides. If all winning workers embed their true locations in sensing reports, their reports correctly record what they sense, and thus $loss = 0$.

C. Adversary Model

The Bayesian attack is the de-facto standard adversary model adopted to measure location privacy since Andrés et al.'s work *geo-indistinguishability* [14]. The adversary, i.e., the platform in this work, is assumed to have side information about a worker's location, knowing, for example, which city it is at. The adversary's side information can be modeled by a prior distribution on $\mathcal{L}_j$, the entire set of w_j's possible locations. $\Pr[l_j]$ ($l_j \in \mathcal{L}_j$) is the probability associated with location l_j. $\Pr[z_j|l_j]$ is the probability that the reported location z_j is converted from l_j, which is assumed to known by the platform as well. Upon observing z_j, the adversary can build a posterior distribution over the inputs, denoted as $\Pr[l_j|z_j]$

$$\Pr[l_j|z_j] = \frac{\Pr[z_j|l_j]\Pr[l_j]}{\sum_{l'_j \in \mathcal{L}_j} \Pr[z_j|l'_j]\Pr[l'_j]} \quad \forall l_j \in \mathcal{L}_j. \qquad (2)$$

Then the platform derives its best guess over w_j's location by looking for the one that produces the largest posterior probability $l_j^* = \arg\max_{l_j \in \mathcal{L}_j} \Pr[l_j|z_j]$. Besides, the platform is assumed working under *semi-honest mode*, i.e., it is trusted to correctly execute protocols of MCS systems but is curious about worker's locations.

D. Auction Model

To motivate workers to participate in MCS, effective incentive mechanisms are indispensable. In this work, we implement the incentive mechanism via a combinational reverse auction. Given a set of sensing tasks announced by the platform, each worker selects a bundle of tasks it is interested in and derives a bid, i.e., the minimum payment it accepts for executing these tasks to compensate its cost.

Denote by c_j as w_j's total cost for sensing tasks in one auction. $c_j = c_j^s + c_j^p \xi_j$ is composed of two parts, w_j's cost for resource consumption, namely sensing cost which is denoted as c_j^s, and cost for privacy leakage, namely privacy cost which is denoted as $c_j^p \xi_j$. The integration of these two makes sure that workers are compensated from both aspects. For any worker w_j, its privacy cost $c_j^p \xi_j$ is positively correlated with its privacy leakage budget ξ_j. Intuitively, the larger privacy disclosed to the platform, the higher it costs to the worker. Therefore, we adopt the natural linear model for privacy cost as in [20] where c_j^p denotes the worker's unit cost of privacy.

We assume that each worker is strategic and aims to maximize its own utility, which is defined as $u_j = p_j - c_j$, where p_j and c_j stand for its payment and cost, respectively. Apparently, if a worker is not selected by the platform, its utility is 0.

III. Design Objectives

In this work, we aim to establish a comprehensive MCS market that exhibits the following desirable properties.

First of all, workers are allowed to choose how much location privacy to disclose to the platform. Following the location privacy modeling in *geo-indistinguishability* [14], we define *ξ_j-privacy*.

Definition 1. ξ_j-Privacy. *A worker w_j achieves ξ_j-*privacy*, if the platform $\mathcal{A}$, who adopts Bayesian attack model, has* $\exp(\xi_j)$ *advantage in inferring w_i's actual location distribution. $\mathcal{A}$'s advantage is defined as*

$$\mathcal{A}_{\mathit{Adv}}[l_j] = \frac{\Pr[l_j|z_j]}{\Pr[l_j]} \le \exp(\xi_j) \qquad \forall \quad l_j \in \mathcal{L}_j \qquad (3)$$

Following [14], ξ_j is specified by a tuple (ϵ_j, r_j), $\xi_j = \epsilon_j / r_j$. r_j is the radius worker w_j is mostly concerned with and ϵ_j is the privacy level it wishes for that radius. The main idea behind this notion is that, for any radius r_j, w_j enjoys ϵ_j-privacy within r_j, i.e., the level of privacy is proportional to the radius. Thus, ξ_j corresponds to the privacy level for one unit of distance.

The platform's advantage is in fact its *posterior knowledge gain*. The goal of privacy protection here is to restrict the information leakage caused by the observation. When w_j sets ξ_j as 0, the platform gains no advantage in inferring w_j's actual location based on the observation over z_j. Note that the lack of leakage does not mean that the worker's location cannot be inferred (it could be inferred by the prior alone), but the observation would not increase the adversary's knowledge. More importantly, the platform's posterior knowledge gain is controlled by w_j through tuning ξ_j. By selecting a proper ξ_j, w_i determines how much advantage the platform can gain.

As the second objective, we aim to achieve a measurable MCS service accuracy at the platform. To protect location privacy from the platform, workers embed obfuscated locations in their reports, and thus inevitably cause *geo-information loss*. Besides, as each worker chooses its obfuscated location in a probabilistic manner (see Section IV-B), $loss$ is in fact a random variable, which brings a great challenge in service accuracy measurement. Instead, we propose to adopt a probabilistic evaluation form.

Definition 2. *(α, β)***-Accuracy**. *The platform provides (α, β)-* accurate *MCS services, if* $\Pr[loss \le \alpha] \ge \beta$*, where $\alpha > 0$ and $\beta \in (0, 1)$.*

Generally, for a given β, a smaller α indicates a better service accuracy. β can be treated as a confidence level for the statement $\Pr[loss \le \alpha]$.

Most of commercialized crowdsensing platforms are typically constrained by a monetary budget B, i.e., the maximum total payment it can afford to reward winning workers in one auction. Thus, it is more practical to take it into account during mechanism design.

Definition 3. Budget Feasibility. *The platform is budget feasible, if* $\sum_{j: w_j \in \mathcal{W}^*} p_j \le B$.

In addition to the above three objectives, we also need to guarantee the following two economic properties that are indispensable for a robust auction-based market.

Definition 4. Truthfulness. *An MCS market is truthful if for any worker w_j, $u_j(c_j, \boldsymbol{b}_{-j}) \geq u_j(b_j, \boldsymbol{b}_{-j})$ where b_j is w_j's submitted bid with $b_j \neq c_j$ and $\boldsymbol{b}_{-j}$ is the bidding profile from other workers except w_j.*

Definition 5. Individual Rationality. *An MCS market is individual rational if each worker w_j has a nonnegative utility $u_j \geq 0$.*

IV. Mechanism Design

A. Problem Formulation

Based on the design objectives discussed in the previous section, we formulate them into the following geo-information loss minimization problem (GLMP)

$$\begin{aligned}
\min: \quad & \alpha \\
\text{s.t.} \quad & \sum_{j:w_j \in \mathcal{W}} x_j p_j \leq B && (4) \\
& \bigcup_{j:w_j \in \mathcal{W}, x_j=1} \mathcal{T}_j \supseteq \mathcal{T} && (5) \\
& \Pr[loss \leq \alpha] \geq \beta && (6) \\
& x_j \in \{0,1\},\ p_j \geq 0,\ \alpha > 0
\end{aligned}$$

Given a fixed confidence level β, the platform aims to minimize the *geo-information loss* of MCS services, while satisfying a series of constraints. Specifically, (4) is due to the *budget feasibility* requirement. (5) states that the platform's sensing task set should be fully covered. (6) is the (α, β)-*accuracy* requirement. x_j is a binary variable. It is equal to 1, if w_j is selected as the winner for sensing tasks; and 0 otherwise. The variable p_j decides the payment to w_j. In addition to constraint (4)-(6), any solution to GLMP should also satisfy some other inherent constraints, including ξ_j-*privacy*, *truthfulness* and *individual rationality*. Due the lack of explicit expressions, we temporarily omit them from the formulation of GLMP.

B. Location Obfuscation Mechanism Design

Solving GLMP is largely hindered by constraint (6), because the variable α is within a probabilistic expression. As far as we know, there is no existing efficient algorithm to directly solve optimization problems in such a form. It drives us to explore other relations among α, β and *loss* in (6), so as to convert it into a form that is easier to handle. As the calculation of $loss$ depends on the location obfuscation mechanism adopted by each worker. In the following, we first introduce our proposed obfuscation mechanism and then provide the corresponding alternative expression for (6).

The mechanism is composed of two procedures, *obfuscated location set generation* and *probabilistic mapping*.

Obfuscated Location Set Generation. For each worker w_j, consider a system of polar coordinates with origin at its true location l_j. w_j determines the unit Δr_j and $\Delta\theta_j$ and evenly divides r_j into $[0, \Delta r_j, 2\Delta r_j, \cdots, r_j]$, and 2π into $[\Delta\theta_j, 2\Delta\theta_j, \cdots, 2\pi]$. Note that r_j is the radius that w_j is most concerned with. Let it be w_j's maximum obfuscation range. Then its obfuscated location set is generated by $\mathcal{Z}_j = \{z_j = (m \cdot \Delta r_j, n \cdot \Delta\theta_j) : m \in [1, \frac{r_j}{\Delta r_j}], n \in [1, \frac{2\pi}{\Delta\theta_j}]\}$, where $(m \cdot \Delta r_j, n \cdot \Delta\theta_j)$ is the polar coordinate of an obfuscated location z_j, with $m \cdot \Delta r_j$ and $n \cdot \Delta\theta_j$ its radius and angle, respectively.

Probabilistic Mapping. Once $\mathcal{Z}_j$ is ready, w_j needs to select among them an element and report it as its location. Our design is motivated by the *exponential mechanism* [22], [23]: for any $z_j \in \mathcal{Z}_j$, its probability of being selected is determined by

$$\Pr[z_j|l_j] = \frac{\exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]}{\sum_{z'_j \in \mathcal{Z}_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z'_j))]}. \quad (7)$$

Apparently, the location z_j with a shorter drift distance $d(l_j, z_j)$ has a higher chance to be chosen. But this advantage diminishes as ξ_j decreases. Particularly, when $\xi_j = 0$, i.e., w_j has zero privacy leakage budget and thus imposes the most strict privacy requirement, all elements in $\mathcal{Z}_j$ have the equal chance to be chosen.

Theorem 1. *With the proposed location obfuscation mechanism, the platform provides (α, β)-accurate MCS services. Given β, then*

$$\alpha = \sqrt{\frac{\sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2}{(1-\beta)}} + \sum_{j:w_j \in \mathcal{W}^*} \mu_j \quad (8)$$

where μ_j and σ_j^2 represent the mean and variance of w_j's drift distance.

Proof. Please refer to Appendix A for the proof. □

C. GLMP Reformulation

Theorem 1 specifies the relation between α and β under the proposed location obfuscation mechanism.

Accordingly, GLMP can be reformulated as

$$\begin{aligned}
\min: \quad & \sqrt{\frac{\sum_{j:w_j \in \mathcal{W}} x_j \sigma_j^2}{(1-\beta)}} + \sum_{j:w_j \in \mathcal{W}} x_j \mu_j && (9) \\
\text{s.t.} \quad & (4), (5) \quad x_j \in \{0,1\},\ p_j \geq 0
\end{aligned}$$

which is referred as the reformulated GLMP for the rest of this paper. Like GLMP, any solution to the reformulated GLMP should also satisfy inherent constraints, including *ξ_j-privacy*, *truthfulness* and *individual rationality*.

Comparing GLMP and its reformulated version, the coefficient $\sigma_j/\sqrt{1-\beta} + \mu_j$ can be viewed as w_j's *geo-information loss* caused to MCS services. While the reformulated GLMP gets rid of the troublesome probabilistic constraint (6), it is still at least NP-hard.

Theorem 2. *The reformulated GLMP is at least NP-hard.*

Essentially, the reformulated GLMP can be degenerated into a conventional *weighted set cover problem*, which has been proved as NP-hard [26]. Thus, the reformulated GLMP is at least NP-hard. Due to the limited space, we omit its formal proof here.

D. Heuristic Algorithm Design

Since the reformulated GLMP is at least NP-hard, it is computationally inefficient to optimally solve it. Instead, we propose a heuristic algorithm to derive a solution to $\boldsymbol{x}$ and $\boldsymbol{p}$. It is worth noting that the elimination of the probabilistic constraint (6) facilitates the design for such a heuristic algorithm, while this is much more complicated by working on GLMP directly.

Our heuristic algorithm is composed of two procedures, *winner selection* and *payment determination*. The first procedure determines the winning worker set, i.e., $\boldsymbol{x}$, while the second one calculates payment for each winner, i.e., $\boldsymbol{p}$.

Algorithm 1 Winner Selection

```
Input: $\mathcal{T}$, $\mathcal{W}$, $\beta$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$
Output: $\boldsymbol{x}$
1: $\mathcal{W}' \leftarrow \{w_j \in \mathcal{W} : \frac{b_j}{|\mathcal{T}_j|} \leq \frac{B}{|\mathcal{T}|}\}$, $\mathcal{W}_0 \leftarrow \emptyset$, $s \leftarrow 1$, $\boldsymbol{x} \leftarrow \mathbf{0}$,
   $\boldsymbol{p} \leftarrow \mathbf{0}$
2: $k \leftarrow \arg\max_{j:w_j \in \mathcal{W}' \setminus \mathcal{W}_0} \frac{g_{j|\mathcal{W}_0}}{b_j}$, $\mathcal{W}_1 \leftarrow w_k$
3: while $\mathcal{T} \neq \emptyset$ and $b_k \leq \frac{B}{2} \times \frac{g_{k|\mathcal{W}_{s-1}}}{S(\mathcal{W}_s)}$ do
4:     $x_k \leftarrow 1$
5:     $\mathcal{W}_s \leftarrow \mathcal{W}_s \cup w_k$, $\mathcal{W}^* \leftarrow \mathcal{W}_s$, $\mathcal{T} \leftarrow \mathcal{T} \setminus \ \mathcal{T}_k$
6:     $s \leftarrow s + 1$
7:     $k \leftarrow \arg\max_{j:w_j \in \mathcal{W}' \setminus \mathcal{W}_{s-1}} \frac{g_{j|\mathcal{W}_{s-1}}}{b_j}$
8: end while
9: Return $\boldsymbol{x}$
```

Winner Selection. As shown in Algorithm 1, once receiving bid profiles from all workers, the platform first creates a set $\mathcal{W}'$ by ruling out any worker whose per-task bid $b_j/|\mathcal{T}_j|$ exceeds the platform's per-task budget $B/|\mathcal{T}|$. Recall from the reformulated GLMP that the platform tends to select workers who execute more tasks and provide high-quality sensing reports, i.e., introduce small drift distances. Thus, we define a parameter called *marginal contribution*

$$g_{j|\mathcal{W}_{j-1}} = \frac{1}{F_j}[G(\mathcal{W}_{j-1} \cup w_j) - G(\mathcal{W}_{j-1})]. \tag{10}$$

$\mathcal{W}_{j-1}$ denotes the set of winning workers selected in the $(j-1)$-th iteration of the while loop in Algorithm 1. $G(\mathcal{W}_{j-1}) = |\cup_{w_j \in \mathcal{W}_{j-1}} \mathcal{T}_j|$ is the number of tasks executed by workers in $\mathcal{W}_{j-1}$. $F_j = \sigma_j/\sqrt{1-\beta} + \mu_j$ is obtained from the reformulated GLMP's objective function, approximating the geo-information loss introduced by w_j. In each iteration, the algorithm selects a worker that produces the largest $g_{j|\mathcal{W}_{s-1}}/b_j$ (line 7). Besides, a winner's bid should also meet the following requirement for the budget feasibility

$$b_j \leq \frac{B}{2} \times \frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}_j)}, \tag{11}$$

where $S(\mathcal{W}_j) = S(\mathcal{W}_{j-1}) + g_{j|\mathcal{W}_{j-1}}$ and $S(\mathcal{W}_0) = 0$. The iteration continues until all tasks are assigned.

Payment Determination. Once winning workers are selected, the remaining job is to determine their payments. In order to achieve *truthfulness*, we follow the idea of *critical payment* [24]. A critical payment p_j for winner w_j is set in a way that w_j wins when bidding lower than p_j, and loses otherwise.

Specifically, for each winning worker w_j, the platform runs Algorithm 1 again based on a different input tuple $\{\mathcal{T}, \mathcal{W} \setminus w_j, \beta\}$ and derives another winner set $\overline{\mathcal{W}}^*$. Since $\mathcal{W} \setminus w_j$ excludes w_j, so does $\overline{\mathcal{W}}^*$. Then for each worker $w_l \in \overline{\mathcal{W}}^*$, which is selected in the l-th iteration in Algorithm 1, the platform finds the highest virtual bid $b^v_{j,l}$ such that w_j can substitute w_l to win (in the l-th iteration), if it bids with $b^v_{j,l}$. It implies that $g_{l|\overline{\mathcal{W}}_{l-1}}/b_l \leq g_{j|\overline{\mathcal{W}}_{l-1}}/b^v_{j,l}$. Together with the winner selection criteria from Algorithm 1, this virtual bid should satisfy

$$b^v_{j,l} = \min\left\{\frac{b_l \times g_{j|\overline{\mathcal{W}}_{l-1}}}{g_{l|\overline{\mathcal{W}}_{l-1}}}, \frac{B}{2} \times \frac{g_{j|\overline{\mathcal{W}}_{l-1}}}{S(\overline{\mathcal{W}}_l)}\right\}. \tag{12}$$

Finally, w_j's payment is set as $p_j = \arg\max_{l \in \overline{\mathcal{W}}^*} b^v_{j,l}$, i.e., the maximum achievable virtual bid from $\overline{\mathcal{W}}^*$.

The following theorem shows that our heuristic algorithm is of polynomial-time complexity. Thus, the reformulated GLMP can be efficiently solved.

Theorem 3. *The computation complexity of our heuristic algorithm is upper bounded by* $\mathcal{O}(M^2N)$.

The above statement can be easily derived according to the heuristic algorithm. Due to the limited space, we omit its formal proof here.

Theorem 4. *Denote by* OPT *the optimal result of the reformulated GLMP, then* $\frac{\alpha}{OPT} \leq \frac{B \cdot \sum_{w_j \in \mathcal{W}^*} \frac{|\mathcal{T}_j|}{c_j}}{2S(\mathcal{W}^*)\min_{j\in[1,M]}\{F_j\}}$, *where* α *is the result obtained via the heuristic algorithm.*

Proof. Please refer to Appendix B for the proof. □

V. Property Analysis

In this section, we provide theoretical analysis over the properties achieved by our mechanism, including ξ_j*-privacy*, *budget feasibility*, *truthfulness* and *individual rationality*. Recall that (α, β)*-accuracy* has been proved to exist in Theorem 1.

To show that ξ_j*-privacy* is guaranteed for each worker w_j, we first give the following lemma.

Lemma 1. *With* w_j*'s location obfuscation mechanism, then* $\frac{1}{\exp(\xi_j)} \leq \frac{\Pr[z_j|l_j]}{\Pr[z_j|l'_j]} \leq \exp(\xi_j)$, *where* $l_j, l'_j \in \mathcal{L}_j$ *are* w_j*'s two arbitrary true locations.*

Proof. Please refer to Appendix C for the proof. □

Lemma 1 says that when w_j is at l_j and l'_j, the ratio between the chances that both of them are mapped to the same obfuscated location z_j is bounded by $[\frac{1}{\exp(\xi_j)}, \exp(\xi_j)]$. When $\xi_j = 0$, $\Pr[z_j|l_j] = \Pr[z_j|l'_j]$, i.e., l_j and l'_j have the same chance to map to z_j. In another word, with the observation of z_j, it's difficult for the platform to determine whether this worker locates at l_j or l'_j. Based on Lemma 1, we are ready to present the privacy protection property achieved by our mechanism.

Theorem 5. *Each winning worker $w_j \in \mathcal{W}^*$ achieves ξ_j-privacy via our mechanism, i.e., $\mathcal{A}_{\mathsf{Adv}}[l_j] \leq \exp(\xi_j)$.*

Proof. Please refer to Appendix D for the proof. □

Theorem 5 indicates that each winning worker has full control of its location privacy leakage to the platform. When it has a harsh privacy requirement, it sets a small ξ_j. Specifically, when $\xi_j = 0$, the platform's *posterior knowledge gain* is 1, i.e., no useful information regarding w_j's true location is explorable from any observation. Regarding the losing workers, as they do not upload any sensing report, no location information will be disclosed.

Recall that the platform's budget is B. To avoid its deficit in hosting sensing tasks, our design has to limit winning workers' total payment. Before discussing if this property holds, we would like to introduce the following lemma, which gives an upper bound to each winner's payment.

Lemma 2. *For a winning worker $w_j \in \mathcal{W}^*$, its payment p_j is upper bounded by $B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$.*

Proof. Please refer to Appendix E for the proof. □

With Lemma 2, we can infer $\sum_{w_j \in \mathcal{W}^*} p_j \leq \sum_{w_j \in \mathcal{W}^*} B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)} = B$, i.e., the total payment to winning workers is confined to the platform's budget B. Thus, the budget feasibility exists.

Theorem 6. *The platform is budget feasible.*

The critical economic properties, including *truthfulness* and *individual rationality*, are also achieved via the proposed mechanism.

Theorem 7. *The MCS market is truthful.*

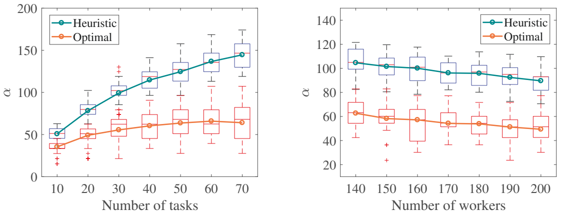
Proof. Please refer to Appendix F for the proof. □

Theorem 8. *The MCS market is individual rational.*

Proof. Please refer to Appendix G for the proof. □

VI. Evaluation

In this section, we provide numerical results on evaluating performances of our mechanism. Real-world dataset retrieved from New York City's 311 platform [25] is adopted. 311 is America's highly popular non-emergency report system allowing people to call in many cities to find information about services, make complaints, or report problems like noise pollution or road damage. In simulations, 167355 data entries from Manhattan area have been extracted. We treat 311 users as sensing workers in MCS and their complaints as sensing reports. Besides, since each complaint is associated with a location coordinate, such information is used to emulate its reporting worker's true location. As a note, 311 dataset has been widely adopted in social/crowd sensing related research. Our code is written in MATLAB on a laptop with 3.4GHz Intel i7 CPU and 16 GB memory.



(a) Impact of task size N. (b) Impact of worker size M.

Fig. 1. Optimality gap of the heuristic algorithm under different settings.

A. Performances of Our Heuristic Algorithm

We first show performances of the proposed heuristic algorithm in terms of its optimality gap and computation efficiency. For this purpose, we compare it with the optimal solution, which is obtained by exhaustive search.

Optimality Gap. The impact of task size N and worker size M are examined in Fig. 1(a) and Fig. 1(b), respectively. The heuristic algorithm produces a higher α than the optimal one. For example in Fig. 1(a), when $N = 20$, $\alpha = 73.46$ for the former, while $\alpha = 48.75$ for the latter. This optimality gap comes from two aspects. Firstly, the heuristic algorithm trades a portion of computation accuracy with computation efficiency, which will be discussed shortly. *More importantly, the exhaustive search does not consider ξ_j-privacy, truthfulness or individual rationality, while these properties have been formally proved to exist in our design through Theorem 5, Theorem 7 and Theorem 8.*

Computation Efficiency. Table 1 compares the computation time for both algorithms under different MCS market sizes. Particularly, under the setting $M = 190$, $N = 100$, it only costs 309.06 ms for the heuristic algorithm to find the solution, while that for the exhaustive search is significantly larger, i.e., 4967.41 ms. The latter is about 16 times the former. Besides, the performance improvement becomes more apparent under a larger market setting. Therefore, our algorithm is suitable for MCS, which typically involves a large number of workers.

B. Privacy Protection

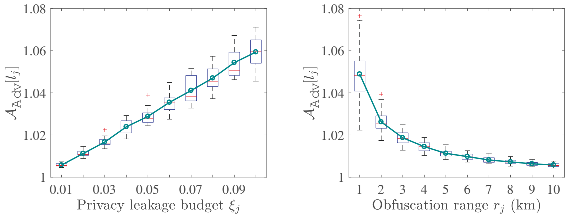
To evaluate the performance of privacy protection, we show the platform's posterior knowledge gain $\mathcal{A}_{\mathsf{Adv}}[l_j]$ toward a worker's true location l_j under the observation of this worker's reported location z_j. An arbitrary worker w_j is randomly selected from the winning worker set and tested.

Impact of Privacy Leakage Budget ξ_j. Fig. 2(a) depicts $\mathcal{A}_{\mathsf{Adv}}[l_j]$ when w_j chooses different ξ_j with a fixed $r_j = 10$km. We observe that the platform's posterior knowledge gain increases as the growth of worker's privacy leakage budget. Specifically, $\mathcal{A}_{\mathsf{Adv}}[l_j] = 1.018$ when $\xi_j = 0.03$, while it reaches 1.043 when $\xi_j = 0.07$. This trend meets the theoretical result derived in Theorem 5. With a smaller ξ_j, the obfuscated location z_j tends to be generated with a more evenly distribution. Thus, the knowledge of z_j provides the platform limited advantage to correctly locate this worker.

MCS market size	$M=40$ $N=20$	$M=50$ $N=30$	$M=70$ $N=50$	$M=100$ $N=80$	$M=140$ $N=80$	$M=170$ $N=90$	$M=180$ $N=90$	$M=190$ $N=100$
Exhaustive search	296.31 ms	592.05 ms	751.05 ms	1226.35 ms	2141.75 ms	2878.43 ms	3384.36 ms	4967.41 ms
Heuristic algorithm	128.69 ms	134.92 ms	156.60 ms	210.37 ms	227.33 ms	256.73 ms	274.61 ms	309.06 ms

TABLE I

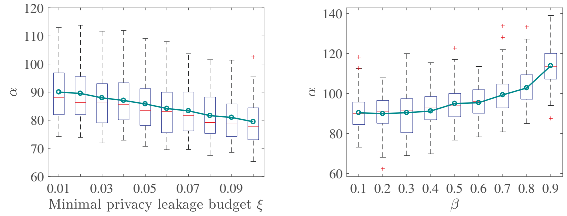
COMPARISON BETWEEN EXHAUSTIVE SEARCH AND THE HEURISTIC ALGORITHM IN TERMS OF COMPUTATION TIME.



(a) Impact of ξ_j.

(b) Impact of r_j.

Fig. 2. Platform's posterior knowledge gain with respect to worker's privacy leakage budget and obfuscation range.



(a) Impact of ξ.

(b) Impact of β.

Fig. 3. Platform's MCS service accuracy with respect to β and workers' privacy leakage budgets.

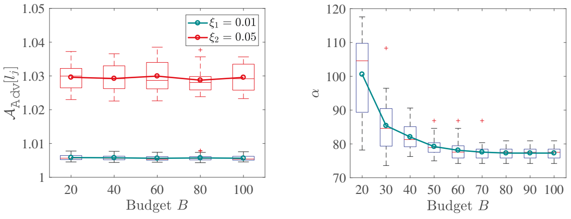
Impact of Obfuscation Range r_j**.** Fig. 2(b) further depicts the impact of obfuscation range r_j to $\mathcal{A}_{\mathsf{Adv}}[l_j]$ under a fixed $\epsilon_j = 0.1$. We notice that $\mathcal{A}_{\mathsf{Adv}}[l_j]$ decreases as the increase of worker's obfuscation range r_j. This is because $\xi_j = \epsilon_j / r_j$ becomes smaller when a larger obfuscation range is chosen, which complies with Theorem 5. Therefore, a worker's location privacy can be better preserved when it chooses a larger obfuscation range.

C. MCS Service Accuracy

We evaluate the service accuracy by examining α with respect to ξ and β, repsectively. Specifically, ξ is defined as $\min_{j:w_j \in \mathcal{W}^*} \xi_j$, i.e., the lowest privacy leakage budget among all winning workers.

Impact of Minimum Privacy Leakage Budget ξ**.** Fig. 3(a) says that a better service accuracy, i.e., a lower α, is achieved when workers select larger privacy leakage budget. Specifically, $\alpha = 90.25$ when $\xi = 0.01$, while it drops to 81.51 when $\xi = 0.09$. Thus, there is a tradeoff relation between the service accuracy and overall privacy protection level. When workers impose strict privacy requirements, it is infeasible for the platform to provide accurate services.

Impact of β**.** Fig. 3(b) depicts the relation between α and β. We observe that α increases as β grows. For instance, $\alpha = 91.42$ when $\beta = 0.3$ and $\alpha = 113.64$ when $\beta = 0.9$. Recall that β is the confidence level for the statement $loss \le \alpha$. Then, when we want to estimate service accuracy with higher confidence, a more conservative and thus a relatively large α will be given.



(a) Impact to $\mathcal{A}_{\mathsf{Adv}}(l_j)$.

(b) Impact to service accuracy α.

Fig. 4. Impact of the platform's budget to MCS market performances.

D. Budget Feasibility

Impact to $\mathcal{A}_{\mathsf{Adv}}[l_j]$**.** Fig. 4(a) examines the platform's advantage in inferring two arbitrary workers' exact locations. These workers adopt privacy leakage budgets 0.01 and 0.05, respectively. We find that the platform's advantage is independent from the budget. This is because each worker determines how much location privacy it leaks to the platform by selecting different privacy leakage budget ξ_j. This value is irrelevant to the platform's budget.

Impact to service accuracy α**.** Fig. 4(b) shows the impact of B to α. We observe that α decreases as B grows. Specifically, $\alpha = 100.26$ when $B = 20$, and it drops to 78.43 when $B = 100$. This is because a large amount of budget allows the platform to recruit workers who provide more accurate sensing reports. As a result, the service accuracy is enhanced.

VII. CONCLUSION

We construct a location privacy trading market for MCS under an auction framework, where the platform provides incentives to motivate workers to complete sensing tasks. Taking into account of budget constraint, service accuracy, and privacy protection, we formulate an optimization problem. To efficiently solve it, a heuristic algorithm is proposed. Formal proofs show that our mechanism achieves *ξ-privacy*, *(α,β)-accuracy*, and *budget feasibility*, which are further validated through extensive simulations based on New York City's 311 platform dataset.

ACKNOWLEDGEMENT

The work of Ming Li was supported by the U.S. National Science Foundation under grants CNS-1566634 and ECCS-1711991. The work of Linke Guo was partially supported by National Science Foundation under grants ECCS-1710996, IIS-1722731 and CNS-1744261.

REFERENCES

[1] C. Cornelius, et al., "Anonysense: privacy-aware people-centric sensing," in *Proceedings of ACM Mobisys*, 2008.

[2] L. Pournajaf, et al., "Spatial task assignment for crowd sensing with cloaked locations," in *Proceedings of the IEEE MDM*, 2014.

[3] I. J. Vergara-Laurens, et al., "Privacy, quality of information, and energy consumption in participatory sensing systems," in *Proceedings of IEEE PerCom*.

[4] L. Wang, et al., "Differential location privacy for sparse mobile crowdsensing," in *Proceedings of IEEE ICDM*, 2016.

[5] L. Wang, et al., "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proceedings of the WWW*, 2017.

[6] "acxiom." [Online]. Available: https://www.acxiom.com/

[7] A. Ghosh and A. Roth, "Selling privacy at auction," *Games and Economic Behavior*, vol. 91, pp. 334–346, May 2015.

[8] L. K. Fleischer and Y.-H. Lyu, "Approximately optimal auctions for selling privacy when costs are correlated with data," in *Proceedings of ACM EC*, 2012.

[9] A. Ghosh and K. Ligett, "Privacy and coordination: computing on databases with endogenous participation," in *Proceedings of ACM EC*, 2013.

[10] A. Ghosh, et al., "Buying private data without verification," in *Proceedings of ACM EC*, 2014.

[11] K. Nissim, et al., "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proceedings of ACM ITCS*, 2014.

[12] W. Wang, et al., "Buying data from privacy-aware individuals: the effect of negative payments," in *Proceedings of WINE*, 2016.

[13] ——, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *Proceedings of the ACM SIGMETRICS*, 2016.

[14] M. E. Andrés, et al., "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of ACM CCS*, 2013.

[15] L. Pournajaf, et al., "Participant privacy in mobile crowd sensing task management: A survey of methods and challenges," *ACM SIGMOD Record*, vol. 44, no. 4, pp. 23–34, May 2016.

[16] X. Wang, et al., "Incentivizing crowdsensing with location-privacy preserving," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6940–6952, October 2017.

[17] Q. Li and G. Cao, "Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error," in *Proceedings on PETS*, 2013.

[18] C. Miao, et al., "Cloud-enabled privacy-preserving truth discovery in crowd sensing systems," in *Proceedings of the ACM Sensys*, 2015.

[19] F. Qiu, et al., "Privacy and quality preserving multimedia data aggregation for participatory sensing systems," *IEEE Transactions on Mobile Computing*, vol. 14, no. 6, pp. 1287–1300, August 2015.

[20] H. Jin, et al., "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proceedings of the ACM MobiHoc*, 2016.

[21] L. Yang, et al., "Crowd-empowered privacy-preserving data aggregation for mobile crowdsensing," in *Proceedings of ACM MobiHoc*, 2018.

[22] Z. Huang and S. Kannan, "The exponential mechanism for social welfare: Private, truthful, and nearly optimal," in *Proceedings of the IEEE FOCS*, 2012.

[23] W. Jin, et al., "DPDA: A Differentially Private Double Auction Scheme for Mobile Crowd Sensing," in *Proceedings of the IEEE CNS*, 2018.

[24] L. Blumrosen and N. Nisan, *Combinatorial auctions*. New York, USA: Cambridge University Press, 2007.

[25] "Nyc311 opendata." [Online]. Available: https://data.cityofnewyork.us/dataset/311-Service-Requests-From-2011/fpz8-jqf4

[26] C. Lund and M. Yannakakis, "On the hardness of approximating minimization problems," *Journal of the ACM*, vol. 41, no. 5, pp. 960–981, 1994.

APPENDIX A
PROOF OF THEOREM 1

As each worker determines its obfuscation mechanism independently, their drift distances are independent with each other as well. From the definition of *geo-information loss*, the mean and variance of *loss* is calculated by $\boldsymbol{E}[loss] = \sum_{j:w_j \in \mathcal{W}^*} \mu_j$ and $\boldsymbol{D}[loss] = \sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2$. Following the *Chebyshev's inequality*, for any nonnegative value a, we have $\Pr[loss - \sum_{j:w_j \in \mathcal{W}^*} \mu_j \geq a] \leq \Pr[|loss - \sum_{j:w_j \in \mathcal{W}^*} \mu_j| \geq a] \leq \frac{1}{a^2} \sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2$. Therefore, $\Pr[loss \leq a + \sum_{j:w_j \in \mathcal{W}^*} \mu_j] \geq 1 - \frac{1}{a^2} \sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2$. Comparing with (6), for a given β, α is calculated by $\alpha = \sqrt{\frac{\sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2}{(1-\beta)}} + \sum_{j:w_j \in \mathcal{W}^*} \mu_j$.

APPENDIX B
PROOF OF THEOREM 4

First of all, we have

$$OPT \geq \min_{j \in [1,M]} \left\{ \sigma_j / \sqrt{1-\beta} + \mu_j \right\} = \min_{j \in [1,M]} \{F_j\}. \quad (13)$$

Besides, via the proposed heuristic algorithm, we obtain a set of winning workers $\mathcal{W}^*$. Then α is expressed as $\alpha = \sqrt{\frac{\sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2}{(1-\beta)}} + \sum_{j:w_j \in \mathcal{W}^*} \mu_j \leq \sum_{w_j \in \mathcal{W}^*} F_j = \sum_{w_j \in \mathcal{W}^*} \frac{(G(\mathcal{W} \cup w_j) - G(\mathcal{W}))}{g_{j|\mathcal{W}_{j-1}}} \leq \sum_{w_j \in \mathcal{W}^*} \frac{|\mathcal{T}_j|}{g_{j|\mathcal{W}_{j-1}}} \leq \sum_{w_j \in \mathcal{W}^*} \frac{B|\mathcal{T}_j|}{2b_j S(\mathcal{W}^*)} = \frac{B}{2S(\mathcal{W}^*)} \sum_{w_j \in \mathcal{W}^*} \frac{|\mathcal{T}_j|}{c_j}$. Note that $\sqrt{\frac{\sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2}{(1-\beta)}} \leq \sum_{j:w_j \in \mathcal{W}^*} \sigma_j / \sqrt{1-\beta}$, as $\sqrt{a^2+b^2} \leq a+b$ $(a, b \geq 0)$.

Meanwhile, following the Algorithm 1 and (11), we have $\frac{g_{1|\mathcal{W}_0}}{b_1} \geq \cdots \geq \frac{g_{j|\mathcal{W}_{j-1}}}{b_j} \geq \cdots \geq \frac{g_{|\mathcal{W}^*||\mathcal{W}_{|\mathcal{W}^*|-1}}}{b_{|\mathcal{W}^*|}} \geq \frac{2S(\mathcal{W}^*)}{B}$. Combining the analysis above, we have $\frac{\alpha}{OPT} \leq \frac{B \cdot \sum_{w_j \in \mathcal{W}^*} \frac{|\mathcal{T}_j|}{c_j}}{2S(\mathcal{W}^*) \min_{j \in [1,M]} \{F_j\}}$ which ends the proof.

APPENDIX C
PROOF OF LEMMA 1

We first prove the correctness for the second inequality where $\Pr[z_j|l_j] \leq \exp(\xi_j) \Pr[z_j|l'_j]$.

For w_j, assume that both l_j and l'_j are mapped to the same obfuscation location z_j via its location obfuscation mechanism. Then $\frac{\Pr[z_j|l_j]}{\Pr[z_j|l'_j]} = \frac{\exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]}{\exp[\frac{\xi_j}{r_j}(r_j - d(l'_j, z_j))]} \times \frac{\sum_{z'_j \in \mathcal{Z}'_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l'_j, z'_j))]}{\sum_{\tilde{z}_j \in \mathcal{Z}_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, \tilde{z}_j))]} = \exp[\frac{\xi_j}{r_j}(d(l'_j, z_j) - d(l_j, z_j))] \leq \exp(\frac{\xi_j}{r_j} r_j) = \exp(\xi_j)$ where $\mathcal{Z}_j$ and $\mathcal{Z}'_j$ stand for the obfuscated location sets w_j generates when it is at l_j and l'_j, respectively. While $\mathcal{Z}_j$ and $\mathcal{Z}'_j$ are different under the same coordinate system, they are identical under their own system of polar coordinates (with origins at l_j and l'_j, respectively) according to the *obfuscated location set generation* procedure; that is, the relative positions between elements in $\mathcal{Z}_j$ and $\mathcal{Z}'_j$ are the same. Thus, $\sum_{z'_j \in \mathcal{Z}'_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l'_j, z'_j))] = \sum_{\tilde{z}_j \in \mathcal{Z}_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, \tilde{z}_j))]$ and the second equality holds. In addition, since $d(l'_j, z_j) \leq r_j$, the inequality also holds.

Following the similar idea, the correctness for the first inequality in the statement can be validated as well.

APPENDIX D
PROOF OF THEOREM 5

The platform's advantage or *posterior knowledge gain* is calculated as $\mathcal{A}_{\mathsf{Adv}}[l_j] = \frac{\Pr[l_j|z_j]}{\Pr[l_j]} = \frac{\Pr[z_j|l_j]}{\Pr[z_j]} = \frac{\Pr[z_j|l_j]}{\sum_{l'_j \in \mathcal{L}_j} \Pr[l'_j] \Pr[z_j|l'_j]} \leq \sum_{l'_j \in \mathcal{L}_j} \frac{\Pr[z_j|l_j]}{\Pr[l'_j] \Pr[z_j|l'_j]} \leq \sum_{l'_j \in \mathcal{L}_j} \frac{\exp(\xi_j) \Pr[z_j|l'_j]}{\Pr[l'_j] \Pr[z_j|l'_j]} = \frac{\exp(\xi_j)}{\sum_{l'_j \in \mathcal{L}_j} \Pr[l'_j]} = \exp(\xi_j)$. The first inequality above is due to the fact that $\frac{1}{a+b} \leq \frac{1}{a} + \frac{1}{b}$ $(a, b > 0)$. The second inequality is derived from Lemma 1. According to Definition 1, each worker w_j achieves ξ_j-privacy in MCS market via our mechanism.

Appendix E
Proof of Lemma 2

Denote by w_j the winning worker selected in the j-th iteration of Algorithm 1, $\overline{\mathcal{W}}^*$ as the winning worker set derived by excluding w_j. Let $r = \arg\max_{l:w_l \in \overline{\mathcal{W}}^*} b^v_{j,l}$, then $p_j = b^v_{j,r}$. Since w_j is not selected in the first $j-1$ iterations from $\mathcal{W}^*$, then $b_j > b^v_{j,l}$ where $l \in [0, j-1]$. It implies $r \geq j$ and thus $\mathcal{W}_{j-1} \subseteq \overline{\mathcal{W}}_{r-1}$. Also, we have $\overline{\mathcal{W}}_{r-1} \cup w_j \subseteq \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$.

When determining payment for w_j, since w_j can substitute w_r to win in the r-th iteration by bidding $b^v_{j,r}$ (and thus p_j), then $p_j \leq \frac{B}{2} \times \frac{g_{j|\overline{\mathcal{W}}_{r-1}}}{S(\overline{\mathcal{W}}_{r-1} \cup w_j)}$. Together with the fact that $\mathcal{W}_{j-1} \subseteq \overline{\mathcal{W}}_{r-1}$, we derive

$$\frac{g_{j|\mathcal{W}_{j-1}}}{p_j} \geq \frac{g_{j|\overline{\mathcal{W}}_{r-1}}}{p_j} \geq \frac{2S(\overline{\mathcal{W}}_{r-1} \cup w_j)}{B}. \tag{14}$$

In the following, we derive the conclusion that $p_j \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$. This discussion should be carried out under all possible cases, $\overline{\mathcal{W}}_{r-1} \cup w_j = \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$ and $\overline{\mathcal{W}}_{r-1} \cup w_j \subset \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$.

For the first case, where $\overline{\mathcal{W}}_{r-1} \cup w_j = \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$, from (14) we have $\frac{g_{j|\mathcal{W}_{j-1}}}{p_j} \geq \frac{2S(\overline{\mathcal{W}}_{r-1} \cup w_j)}{B} = \frac{2S(\overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*)}{B} \geq \frac{S(\mathcal{W}^*)}{B}$ and thus $p_j \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$.

For the second case, where $\overline{\mathcal{W}}_{r-1} \cup w_j \subset \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$, we plan to derive the conclusion via the contradiction method. Specifically, we assume $p_j > B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$. Besides, denote by $\mathbb{W}_1 = \overline{\mathcal{W}}_{r-1} \cup w_j$ and $\mathbb{W}_2 = \overline{\mathcal{W}}_{r-1} \cup \mathcal{W}^*$ for expression simplicity.

Let $r' = \arg\max_{t:w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} \{\frac{g_{t|\mathbb{W}_1}}{b_t}\}$, then

$$\begin{aligned}\frac{S(\mathbb{W}_2) - S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} &\leq \frac{g_{r'|\mathbb{W}_1}}{b_{r'}} \leq \frac{g_{r|\mathbb{W}_1}}{b_r} \leq \frac{g_{r|\overline{\mathcal{W}}_{r-1}}}{b_r} \\ &\leq \frac{g_{j|\overline{\mathcal{W}}_{r-1}}}{p_j} \leq \frac{g_{j|\mathcal{W}_{j-1}}}{p_j} < \frac{S(\mathcal{W}^*)}{B}. \end{aligned} \tag{15}$$

The first inequality is also derived from a contradiction point of view. Assuming $\frac{S(\mathbb{W}_2)-S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} > \frac{g_{r'|\mathbb{W}_1}}{b_{r'}}$, then $\frac{S(\mathbb{W}_2)-S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} > \frac{g_{t|\mathbb{W}_1}}{b_t}$ for $w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1$. Adding up these inequalities and applying some simple transformations, then $\frac{S(\mathbb{W}_2)-S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} > \frac{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} g_{t|\mathbb{W}_1}}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t}$ and thus $S(\mathbb{W}_2) - S(\mathbb{W}_1) > \sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} g_{t|\mathbb{W}_1}$, which contradicts with the fact $S(\mathbb{W}_2) - S(\mathbb{W}_1) \leq \sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} g_{t|\mathbb{W}_1}$ implied by (10). Therefore, the first inequality of (15) must hold. Its last inequality directly comes from the assumption $p_j > B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$.

From the winner selection rule and (11), we have $\frac{g_{1|\mathcal{W}_0}}{b_1} \geq \cdots \geq \frac{g_{j|\mathcal{W}_{j-1}}}{b_j} \geq \cdots \geq \frac{g_{|\mathcal{W}^*||\mathcal{W}_{|\mathcal{W}^*|-1}}}{b_{|\mathcal{W}^*|}} \geq \frac{2S(\mathcal{W}^*)}{B}$. Thus, $\sum_{j:w_j \in \mathcal{W}^*} b_j \leq \sum_{j:w_j \in \mathcal{W}^*} \frac{B}{2} \times \frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)} = \frac{B}{2}$, and thus $\sum_{t:w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t \leq \frac{B}{2}$. Together with (15), $\frac{2(S(\mathcal{W}^*)-S(\mathbb{W}_1))}{B} \leq \frac{S(\mathbb{W}_2)-S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} \leq \frac{S(\mathcal{W}^*)}{B}$, from which we have $S(\mathcal{W}^*) \leq 2S(\mathbb{W}_1)$. Integrating it into (14), we have $p_j \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{2S(\overline{\mathcal{W}}_{r-1} \cup w_j))} = B\frac{g_{j|\mathcal{W}_{j-1}}}{2S(\mathbb{W}_1)} \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$ which contradicts with the assumption that $p_j > B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$.

According to the discussion above, we conclude that $p_j \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$.

Appendix F
Proof of Theorem 7

Suppose worker w_j bids b_j other than its truthful cost c_j. We first consider the scenario where $b_j > c_j$.

- *Case 1:* w_j wins with both c_j and b_j. According to the payment policy, a winning worker's payment is independent to its bid. Thus, in either case it receives the same payment p_j. Therefore, $u_j(c_j, \boldsymbol{b}_{-j}) = p_j - c_j = u_j(b_j, \boldsymbol{b}_{-j})$.
- *Case 2:* w_j wins with c_j but loses with b_j. Therefore, $u_j(c_j, \boldsymbol{b}_{-j}) > u_j(b_j, \boldsymbol{b}_{-j}) = 0$.
- *Case 3:* w_j loses with c_j but wins with b_j. It implies $\frac{g_{j|\mathcal{W}_{s-1}}}{b_j} > \frac{g_{j|\mathcal{W}_{s-1}}}{c_j}$ and thus $c_j > b_j$, which contradicts with the statement $b_j > c_j$. Therefore, this case will not happen.
- *Case 4:* w_j loses with both c_j and b_j. Then $u_j(c_j, \boldsymbol{b_{-j}}) = u_j(b_j, \boldsymbol{b_{-j}}) = 0$.

From the discussion above, $u_j(c_j, \boldsymbol{b}_{-j}) \geq u_j(b_j, \boldsymbol{b}_{-j})$ when $b_j > c_j$. The proof is similar for the scenario where $b_j < c_j$, which is omitted due to space limit. According to Definition 4, we derive the conclusion.

Appendix G
Proof of Theorem 8

For any winner $w_j \in \mathcal{W}^*$, if we can show that $c_j < b^v_{j,l}$ for a certain $w_l \in \overline{\mathcal{W}}^*$, then $c_j < b^v_{j,l} \leq p_j$ and thus the theorem exists.

For this purpose, we identify this worker $w_{l(j)}$ as the one selected in the l-th iteration for the payment determination (and thus the l-th winner of $\overline{\mathcal{W}}^*$) and also selected in the j-th iteration for winner determination (and thus the j-th winner of $\mathcal{W}^*$). Then $\overline{\mathcal{W}}_{l(j)-1} = \mathcal{W}_{j-1}$, and accordingly,

$$\begin{aligned} b_j &\leq \frac{B}{2} \cdot \frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}_j)} = \frac{B}{2} \cdot \frac{g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{S(\overline{\mathcal{W}}_{l(j)-1}) + g_{j|\overline{\mathcal{W}}_{l(j)-1}}} \\ &\leq \frac{B}{2} \cdot \frac{g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{S(\overline{\mathcal{W}}_{l(j)-1}) + g_{l(j)|\overline{\mathcal{W}}_{l(j)-1}}} = \frac{B}{2} \cdot \frac{g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{S(\overline{\mathcal{W}}_{l(j)})}. \end{aligned} \tag{16}$$

From the assumption of $w_{l(j)}$, it can be inferred that it is selected in a later order than w_j in the winner selection procedure, and thus $l > j$. Therefore, $g_{j|\overline{\mathcal{W}}_{l(j)-1}} \geq g_{l(j)|\overline{\mathcal{W}}_{l(j)-1}}$, which explains the second inequality above. Due to the similar reason, we have $\frac{g_{l(j)|\mathcal{W}_{j-1}}}{b_{l(j)}} \leq \frac{g_{j|\mathcal{W}_{j-1}}}{b_j}$, and thus

$$b_j \leq \frac{b_{l(j)} \times g_{j|\mathcal{W}_{j-1}}}{g_{l(j)|\mathcal{W}_{j-1}}} = \frac{b_{l(j)} \times g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{g_{l(j)|\overline{\mathcal{W}}_{l(j)-1}}}. \tag{17}$$

Meanwhile, according to the payment rule of Algorithm 1, $b^v_{j,l(j)} = \min\left\{\frac{b_{l(j)} \times g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{g_{l(j)|\overline{\mathcal{W}}_{l(j)-1}}}, \frac{B}{2} \times \frac{g_{j|\overline{\mathcal{W}}_{l(j)-1}}}{S(\overline{\mathcal{W}}_{l(j)})}\right\}$. Together with (16) and (17), we have $b_j \leq b^v_{j,l(j)}$. Since Theorem 7 states that $b_j = c_j$, then $c_j = b_j \leq b^v_{j,l(j)} \leq \arg\max_{k:w_k \in \overline{\mathcal{W}}^*} b^v_{j,k} = p_j$. According to Definition 5, we derive the conclusion.