ULPT: A User-Centric Location Privacy Trading Framework for Mobile Crowd Sensing

Wenqiang Jin, *Student Member, IEEE,* Mingyan Xiao, *Student Member, IEEE,* Linke Guo, *Member, IEEE,* Lei Yang, *Member, IEEE,* and Ming Li *Member, IEEE*

Abstract—Mobile crowd sensing (MCS) arises as a promising data collection paradigm that leverages the power of ubiquitous mobile devices to acquire rich information regarding their surrounding environment. In many location-based sensing tasks, workers are required to associate their sensing reports with corresponding geographic coordinates. Such information leaves a trail of worker's historical location record which thus poses a severe threat to their location privacy. On the other hand, individual workers may perceive location privacy differently. Instead of following conventional solutions that aim to perfectly hide user privacy, this paper adopts a novel alternative approach. A **u**ser-centric **l**ocation **p**rivacy **t**rading framework, called ULPT, is constructed to facilitate location privacy trading between workers and the platform. Each worker can decide how much location privacy to disclose to the platform in an MCS task based on its own location privacy leakage budget ξ. The higher ξ is, the more privacy its reported location discloses. Accordingly, it receives higher payment from the platform as compensation. Besides, ULPT enables the platform to select a suitable set of winning workers to achieve desirable MCS service accuracy while taking into account of its budget limit and worker privacy requirements. For this purpose, a heuristic algorithm is devised with a bounded optimality gap. As formally proved in this manuscript, ULPT guarantees a series of nice properties, including ξ-*privacy*, (α, β)-*accuracy*, *budget feasibility*. Moreover, both rigorous theoretical analysis and extensive simulations are conducted to evaluate tradeoffs among these three.

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

1 Introduction

1.1 Motivation

MOBILE crowd sensing (MCS) emerges as a new trend that harnesses a plethora of mobile devices with their on-board sensors (e.g., camera, gyroscope, 3D accelerometer, and GPS) to collect diverse data from the surrounding environment. Nowadays, quite a few MCS systems [1], [28], [29] have been deployed, with their focuses on noise mapping, smart transportation, road surface monitoring, indoor floor plan reconstruction, health care, and many others. Since sensing tasks are generally location-dependent, participating individuals, i.e., workers, are required to embed their locations into sensing reports, which inevitably raises privacy concerns when the sensing data are released to third-party entities.

There are some prior works on protecting worker's location privacy in MCS [30]–[34]. Conventional approaches, such as cloaking and k-anonymity, aim to perfectly protect location privacy for every worker. In fact, individuals may perceive their privacy differently; some may impose stringent requirements over privacy leakage, while some others accept monetary reward in the trade of their personal data. Under this observation, some initial efforts have been devoted to constructing a data trading market that facilitates individuals to sell their privacy in a controllable manner [36]–[40]. In these works, there exists a trustworthy "agent" for individual users to interact with data buyers. In other words, users do not have control over their own data. In this concern, Wang et al. [41], [42] proposed to have each user independently decide how much privacy to sell. A user chooses to add a certain amount of noise to its original data in accordance with its privacy requirement. Clearly, adding noise would degrade the accuracy of MCS services. Yet, there is a lack of a theoretical framework that analyzes privacy-accuracy tradeoffs. Yang et al. [55] recently employed an auction approach that takes into account service accuracy during privacy trading. Still, users cannot customize how much noise to add to their reported data; instead, the noise distribution is determined by the agent.

Under these observations, in this paper we propose a **u**ser-centric **l**ocation **p**rivacy **t**rading framework (ULPT) featured two functions. (i) Workers take full control of their own privacy and determine how much location privacy to disclose to the platform by tuning the amount of noise embedded in the reported geographic coordinates; and (ii) the platform can decide which subset of workers to select by taking into account of target MCS service accuracy and budget limit. Nonetheless, this is not an easy task with challenges mainly from the following aspects.

First of all, it is crucial for a worker to quantify its privacy loss by adding different amounts of noise to its location. With this basis, the worker is able to estimate its privacy cost and thus the minimum acceptable compensation to participate in an MCS task. Typically, the more accurate a worker reports its location, the higher privacy loss it experiences, and thus the higher compensation it demands. On the other hand, the platform is usually limited

- *W. Jin, M. Xiao and M. Li are with the Computer Science and Engineering Department, the University of Texas at Arlington.*
- *L. Guo is with the Electrical and Computer Engineering Department, Clemson University.*
- *L. Yang is with the Computer Science and Engineering Department, University of Nevada Reno.*

by its monetary budget in practice. Such a constraint can significantly impact trading outcomes. For example, even sufficient workers are willing to sell their location privacy, the platform may be short of budget to afford such cost. Therefore, as a second challenge, *budget feasibility* needs to be counted in the mechanism design. Besides, in our scheme the platform has no control regarding how much a worker obfuscates its reported location. Service accuracy is achieved by proper worker selection, jointly considering their sensing capacity and privacy requirements, in addition to the platform's own budget feasibility.

1.2 Summary of Main Contributions

To model the interaction between workers and the platform, we adopt the framework of the reverse auction, where the platform acts as an auctioneer that holds a set of tasks and elicits workers to participate by paying them for both sensing efforts and potential location privacy leakage caused by reported data. A *geo-information loss* minimization problem is formulated that identifies a proper set of workers to achieve the best service accuracy under the platform's given budget. Our main contributions are summarized as follows.

- ξ-*Privacy.* We propose a location privacy quantization method by leveraging the notion of geo-indistinguishability. Each worker has its own *location privacy leakage budget* ξ, which implies the maximal amount of advantage an adversary can gain in inferring its actual location from the reported data. The larger ξ is, the loose requirement a worker has. It also indicates that the worker is more willing to sell its location privacy to the platform.
- (α, β)-*Service Accuracy.* Under the proposed location privacy quantization method, a novel location obfuscation mechanism is devised, which allows each worker to choose its obfuscated location to report in a probabilistic manner. In order to precisely measure MCS service accuracy, we define an (α, β)-*accuracy*, $\Pr[loss \leq \alpha] \geq \beta$, where $loss$ is the service degradation caused by workers' reported obfuscated locations and β is a confidence level. The introduction of (α, β)-*accuracy* assists to derive explicit expressions of tradeoff relations among accuracy, privacy, and budget.
- *Auction based Privacy Trading.* To facilitate the trading of location privacy, we construct an auction that allocates sensing tasks to a set of workers that minimize the overall *geo-information loss* while taking account of various constraints including *privacy*, *accuracy*, *budget feasibility*, *truthfulness* and *individual rationality*. We show that it is NP-hard to find its optimal solution. Thus, a heuristic algorithm, which delivers bounded sub-optimal solutions, is developed with computation complexity bounded by $\mathcal{O}(M^2N)$.
- *Tradeoff Relations among Privacy, Accuracy, and Budget Feasibility.* In addition to achieving all these properties via our design, we further theoretically analyze their tradeoff relations through Theorem 9 to Theorem 11. These theoretical results provide insights for designing desirable MCS markets. It is worth noting that such a comprehensive discussion is missing from previous literature. We also validate the theoretical results via extensive simulations using real-world datasets.

2 Related Work

2.1 Protecting Location Privacy in MCS

Protecting location privacy in MCS has attracted increasing attention. Existing schemes mainly fall into three categories, cloaking [30]–[32], k-anonymity [45], and location obfuscation [2], [33], [34]. However, these schemes are in need of a trusted third party, which typically does not exist in MCS systems. Moreover, they fail to consider scenarios where the adversary has some prior knowledge about a worker's real location. Recently, the notation of *geo-indistinguishability* [43], a novel combination between differential privacy and location privacy introduced by Andrés et al., has been employed. Wang et al. [33] proposed a *geo-indistinguishability* based location preserving mechanism to defend the honest-but-curious platform; a linear programming problem is formulated to minimize the data uncertainty. Following a similar idea, they further study location privacy protection during sensing task allocation in MCS [2], [34] and spectrum sensing [3]. Wang et al. [10] developed a novel location obfuscation mechanism for the MCS under the framework of differential privacy and distortion privacy. Gao et al. [11] and Yi et al. [12] devised location privacy schemes based on crypto primitives. Domi et al. [13] generated a set of dummy locations to camouflage a worker's true location. All the above works treat location privacy from different workers equally. In fact, individuals may perceive differently towards their privacy. Under this observation, in this work we take an alternative approach for privacy protection. Workers are provided with flexibility in determining their privacy level and selling their location information in trade of monetary rewards.

2.2 Protecting Data Privacy in MCS

Data privacy has also been studied in the context of MCS, e.g., [46]. They aim to protect the worker's reported data from the platform, since individual data can potentially cause leakages of sensitive information regarding their reporters, such as locations, trajectories, habits, preferences and so on. Techniques, such as cryptographic multi-party computation [47], [48], data perturbation [4], [5], [49]–[51] and anonymization [52], [53] have been employed. Recently, Jin et al. [54] developed a privacy-preserving data aggregation scheme based on incentive mechanism considering workers' reliability and privacy cost. Note that its objective is to protect worker data privacy from outsiders. And, the platform is assumed trustworthy. Instead, in this work we aim to protect worker's privacy from the platform. [55] also adopts the privacy trading framework, where workers add noise to their original data and get paid accordingly, they cannot customize how much noise to add; the noise distribution is determined by the platform. A similar problem is discussed in [56].Yang et al. [4] model the MCS market as Stackelberg games, where the platform determines the payment policies and workers choose how much noises to

add into their data. However, they assume the workers' data are mutually correlated and fit one statistical structure. While, in this work, we do not have such hard constraints over the worker's private data. Zhang et al. [56] and Sun et al. [8] exploited the contract theory to protect worker's private sensing data. However, workers are required to choose their privacy-preserving levels from a limited number of contracts designed by the platform. In contrast, we do not have such hard restrictions in our design. Workers can freely select their privacy protection level via the privacy leakage budget. Xiong et al. [9] proposed a personalized privacy protection framework based on game theory and data encryption. Rather than data privacy, we focus on location privacy. More importantly, we are combating a more powerful attacker which already has some prior knowledge regarding the worker's location. Thus, our objective is to prevent the attacker from gaining extra information on a worker's location during observations. Accordingly, our technique and design rationality significantly deviate from the existing ones.

2.3 Selling Privacy

Treating user privacy as commodities, Ghosh and Roth [36] are among the first to lay a theoretical foundation for selling private data. Their mechanism asks users to report their cost for the use of their private data to estimate certain statistics, then selects and pays some of them accordingly based on their stated cost. Since user's cost and their personal data are correlated, this mechanism is vulnerable to unexpected data privacy leakage by exploring this correlation. To resolve this issue, Fleischer and Lyu [37] further looked into privacy protection over individual payments. [6], [36], [38]–[40] also fall into this line of research. However, these works assume the existence of an "agent" to sell user's data. This agent is assumed trustworthy and is willing to protect user privacy. Very recent works [41], [42] propose private data trading that users take full control of their own privacy. Nonetheless, they adopt game-theoretic models, which, however, may end up with an inefficient equilibrium; the accuracy of data aggregation is not guaranteed.

3 System Overview

3.1 MCS Systems

We consider a general MCS system consisting of a platform that hosts a set of N location-dependent sensing tasks $\mathcal{T} = \{\tau_1, \cdots, \tau_i, \cdots, \tau_N\}$, and a set of participating workers $\mathcal{W} = \{w_1, \cdots, w_j, \cdots, w_M\}$. Workers are required to report their location coordinates along with their sensory data. However, directly reporting locations impairs workers' privacy and thus discourages their participation. Therefore, to protect their location privacy from the platform, workers are allowed to embed obfuscated locations z_j in the reports instead of their genuine locations l_j. Different workers may value their location privacy differently. The selection of z_j depends on their self-determined privacy budget ξ_j, which represents the maximum amount of privacy w_j is willing to disclose. Generally, the lower ξ_j is, the more stringent requirement it imposes over its location privacy. As a result, less useful locational information z_j leaks to the platform that can be leveraged to infer l_j. However, obfuscated locations affect the accuracy of data aggregation. The platform needs to carefully determine which subset of workers to select so as to ensure accuracy-guaranteed MCS services. The framework of our proposed MCS market ULPT is illustrated in Fig. 1, and its workflow is summarized as follows.

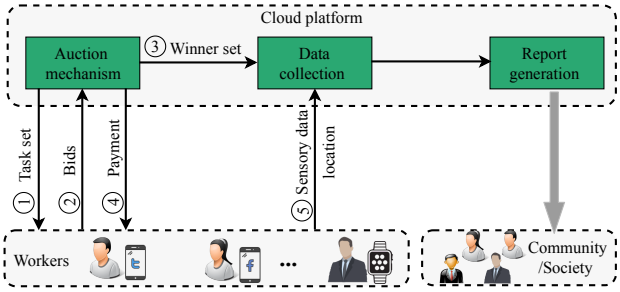


Fig. 1: The workflow of ULPT.

- To incentive worker participation, the platform implements a reverse auction by firstly announcing a set of sensing tasks $\mathcal{T}$ (step ①).
- Then, each worker $w_j \in \mathcal{W}$ submits its interested sensing tasks $\mathcal{T}_j$ and bid b_j to the platform (step ②), where the bid reflects w_j's the minimum payment it accepts to compensate its costs c_j. Following [36], [54], [55], workers take into account the cost c_j^s on resource consumption during task sensing, in addition to privacy loss $c_j^p \xi_j$ when estimating costs, where c_j^p denotes the unit cost of privacy. Therefore, $c_j = c_j^s + c_j^p \xi_j$.
- Following certain criterion (see Section 5.1), the platform determines a winner set $\mathcal{W}^*$ (step ③), i.e., the workers to execute sensing tasks, and their payments $\boldsymbol{p}$ (step ④).
- 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 submits them to the platform (step ⑤). Finally, the platform aggregates over collected sensing reports, derives sensing results, and publishes them to the community or sends back to task requestors.

We summarize all symbols and their definitions involved in this paper in Table 1.

3.2 Adversary Model

As tasks are disseminated by the platform, it knows which geographic region a task pertains to. Such side information renders the platform a much powerful adversary to infer a worker's privacy from. We model the platform as the Bayesian attacker. The Bayesian attack is the de-facto standard adversary model adopted to measure location privacy since Andrés et al.'s work *geo-indistinguishability* [43]. The adversary, i.e., the platform in this work, has the side information as a prior distribution on $\mathcal{L}_j$, which is 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

TABLE 1: Notations.

Symbols	Definition
$\mathcal{T}$	Sensing task set
$\mathcal{W}$	Worker set
τ_i	Task i
w_j	Worker j
$\mathcal{T}_j$	w_j's interested task set
z_j	Obfuscated location
$\mathcal{Z}_j$	Obfuscated location set
l_j	Genuine location
$\mathcal{L}_j$	w_j's possible location set
ξ_j	Privacy budget
r_j	Obfuscation range
ϵ_j	Privacy level
x_j	Winner selection indicator
b_j	Bid price
p_j	Payment
c_j^s	Cost of resource consumption
c_j^p	Cost of privacy loss
u_j	w_j's utility
$\mathcal{W}^*$	Winning worker set
α	Service accuracy
β	Confidence level
B	Platform's budget
$\boldsymbol{b}_{-j}$	Bidding profile from other workers except w_j
Δr_j	Radius unit
$\Delta\theta_j$	Angle unit
$d(l_j, z_j)$	Drift distance
$loss$	Geo-information loss
μ_j	Mean value of drift distance
σ_j	Variance value of drift distance
$g_{j\|\mathcal{W}_{j-1}}$	w_j's marginal contribution
$S(\mathcal{W}_j)$	Sum of worker's marginal contributions in $\mathcal{W}_j$
$\mathcal{A}_{\mathsf{Adv}}[l_j]$	Adversary's gained advantage

l_j, which is assumed to known by the platform as well via long-term observation. Upon observing z_j, the adversary can build a posterior distribution over the inputs, denoted as $\Pr[l_j|z_j]$

$$\Pr[l_j|z_j] = \frac{\Pr[z_j|l_j]\Pr[l_j]}{\sum_{l'_j\in\mathcal{L}_j}\Pr[z_j|l'_j]\Pr[l'_j]} \quad \forall l_j \in \mathcal{L}_j. \tag{1}$$

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 MCS protocols but is curious about worker's locations.

3.3 Privacy Quantization Model

Workers are allowed to determine how much location privacy to disclose to the platform. Following the *geo-indistinguishability* [43] model, we define the quantization of location privacy as follows.

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. We define $\mathcal{A}$'s advantage as*

$$\mathcal{A}_{\mathsf{Adv}}[l_j] = \frac{\Pr[l_j|z_j]}{\Pr[l_j]} \leq \exp(\xi_j) \qquad \forall \quad l_j \in \mathcal{L}_j \tag{2}$$

where $\xi_j \in [0, +\infty)$ is w_j's location privacy leakage budget determined by w_j.

Motivated by [43], the privacy budget ξ_j is determined by a tuple (ϵ_j, r_j), $\xi_j = \epsilon_j/r_j$. Specifically, r_j is the radius of worker w_j's mostly concerned obfuscation area. ϵ_j is the privacy level it desired for that radius. The rationality behind this notion is that, for an given 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 essentially its *posterior knowledge gain*. In this work, the objective of privacy protection is to restrict the information leakage caused by the observations of worker's obfuscated location z_j. When w_j sets its privacy budget ξ_j as 0, from the definition above, the platform has no gained advantage in inferring w_j's genuine location based on its observations. Noteworthy, lack of leakage does not necessarily means that l_j cannot be inferred (it could be inferred from the prior alone), instead that the adversary's knowledge does not increase due to the observation. More importantly, the platform's posterior knowledge gain is fully determined by w_j through tuning ξ_j. By selecting a proper ξ_j, w_j controls how much advantage the platform can gain over w_j's location distribution from its observation.

4 Design Objectives

To protect the location privacy from the platform, a worker w_j adopts a location obfuscation mechanism (see Section 5.2) to generate an obfuscated location z_j from its genuine location l_j. Then, w_j reports z_j together with its sensing data. Since sensing tasks are location dependent, if a worker's reported location deviates from the original one, where sensing data are actually acquired, its sensing report experiences some information loss, so does the aggregation result derived at the platform. Since this information loss is the result of worker's location obfuscations, we define it as *geo-information loss*

$$loss = \sum_{j:w_j\in\mathcal{W}^*} d(l_j, z_j) \tag{3}$$

where $d(l_j, z_j)$ represents w_j's drift distance between its genuine location l_j and the obfuscated one z_j. *Geo-information loss* is the summation of all winning workers' drift distance. The larger value of $loss$, the more "noises" are added to obfuscate the worker's locational information and thus more coarse z_j will be. On the one hand, it implies that workers' location privacy are well preserved. However, the MCS platform will suffer from poorer service accuracy. If all selected workers report their genuine locations, then $loss = 0$ and implies the reports correctly record what they sense with no accuracy degradation.

As each worker chooses its obfuscated location in a probabilistic manner (see Section 5.2), $loss$ is in fact a random variable, which brings a great challenge in service accuracy measurement. Instead, we adopt a probabilistic evaluation form to achieve a measurable MCS service accuracy at the platform.

Definition 2. **(α, β)-Accuracy**. *The platform provides (α, β)-*accurate *MCS services, if $\Pr[loss \leq \alpha] \geq \beta$ 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 $\Pr[loss \leq \alpha]$.

Meanwhile, commercialized MCS platforms are typically budget constrained, i.e., it holds a maximum amount of monetary rewards B for compensating the selected sensing workers in one auction. Thus, it is more practical to discuss the privacy and service accuracy properties under this constraint.

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

In addition to the above two objectives for a desirable MCS market, we also aim to guarantee the following critical economic properties.

Following the conventional auctions, in this work workers are assumed to be strategic that can manipulate its own bids 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, workers that are not selected by the platform will have no costs and payments at all, thus its utility is 0. However, one of our objectives is to design a *truthful* incentive mechanism defined in Definition 4.

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.

Truthfulness is crucial in an auction. Workers may be strategic in a sense to manipulate their bids to win an auction. Truthfulness ensures that the best strategy for a worker is to bid following its true cost. As the platform does not know worker's costs, including both sensing costs and privacy costs, truthfulness is an effective approach to prevent market manipulation which affects interests from other workers and the platform. Besides, truthfulness simplifies the strategic decision process for all workers, as their true costs are the best strategies [63].

Moreover, to guarantee that worker's costs are well compensated, such that they are willing to participate in MCS, another desirable property is *individual rationality*.

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

5 Mechanism Design

5.1 Problem Formulation

Given a fixed confidence level β, the platform aims to minimize the *geo-information loss* of MCS services, while satisfying a series of design objectives discussed in the previous section. Thus, we formulate them into the following geo-information loss minimization problem (GLMP).

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

Specifically, (4) is the *budget feasibility* constraint. (5) ensures that the required sensing task set can be fulfilled by the selected workers. (6) is the (α, β)-*accuracy* requirement. x_j is a binary indicator. $x_j = 1$ represents the worker w_j is selected as a winner for executing the sensing tasks, while $x_j = 0$ states the worker loses the auction. In addition to constraint (4)-(6), the solutions to GLMP should also satisfy some 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.

5.2 Location Obfuscation Mechanism Design

Directly solving GLMP using the conventional optimization tools can be quite challenging, because the constraint (6) is essentially a probabilistic expression. It motivates us to further investigate the close relations among α, β and $loss$ in (6), so as to convert it into a form that is easier to handle. Since $loss$ depends on workers' location obfuscation mechanisms, we first introduce the mechanism design of location obfuscations and then derive the corresponding alternative expression for (6).

The location obfuscation mechanism composes two procedures, i.e., *obfuscated location set generation* and *probabilistic mapping*.

Obfuscated Location Set Generation. For each worker w_j, the obfuscated locations are generated from a polar coordinate system with its genuine location l_j as the origin. Let r_j be w_j's maximum obfuscation range. Within w_j's radius of concerned area, it evenly divides r_j into $[0, \Delta r_j, 2\Delta r_j, \cdots, r_j]$, and 2π into $[\Delta\theta_j, 2\Delta\theta_j, \cdots, 2\pi]$ with its customized unit Δr_j and $\Delta\theta_j$ as shown in Fig. 2. Then its obfuscated location set is generated as $\mathcal{Z}_j = \{z_j = (m \cdot \Delta r_j, n \cdot \Delta\theta_j) : m \in [1, \frac{r_j}{\Delta r_j}], n \in [1, \frac{2\pi}{\Delta\theta_j}]\}$, 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.

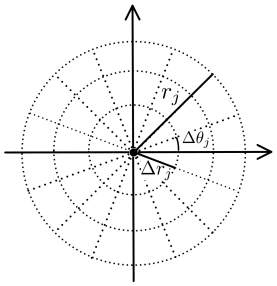


Fig. 2: Obfuscated Location Set Generation

Probabilistic Mapping. Having the location set $\mathcal{Z}_j$, the worker w_j needs to select its obfuscated location z_j among them to substitute l_j in its sensing reports. Motivated by [57], we design a location probabilistic mapping scheme by leveraging the *exponential mechanism*: 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 poses the highest 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*

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

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

Proof. Since workers determine their location obfuscation mechanism independently, the formed obfuscation set $\mathcal{Z}_j$ and probabilistic mapping distribution (7) are different so as their drift distances. From the *geo-information loss* (3) definition, the mean and variance of $loss$ can be calculated as $\boldsymbol{E}[loss] = \sum_{j:w_j \in \mathcal{W}^*} \mu_j$ and $\boldsymbol{D}[loss] = \sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2$. Applying 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 $\alpha = \sqrt{\frac{\sum_{j:w_j \in \mathcal{W}^*} \sigma_j^2}{(1-\beta)}} + \sum_{j:w_j \in \mathcal{W}^*} \mu_j$. □

5.3 GLMP Reformulation

Theorem 1 specifies the relation between α and β under the proposed location obfuscation mechanism. Therefore, GLMP can be reformulated as

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

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 coefficients $\sigma_j/\sqrt{1-\beta}$ and μ_j can be viewed as w_j's *geo-information loss* caused to MCS services. Besides, it can be inferred from the location obfuscation mechanism that, for a given r_j, a smaller ξ_j leads to a larger $\sigma_j/\sqrt{1-\beta}+\mu_j$.While the reformulated GLMP gets rid of the troublesome probabilistic constraint (6), it is still at least NP-hard to optimally solve.

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

Proof. We first degenerate the problem into a special case without considering ξ_j-*privacy*, *truthfulness* or *individual rationality*. Besides, each worker w_j designs its location obfuscation mechanism in a way such that $\sigma_j = 0$. Then, the degenerated problem becomes a conventional *weighted set cover problem*. Specifically, given a set of elements, called the universe, and a collection of weighted sets whose union equals the universe, *weighted set cover problem* is to identify the smallest weighted subcollection of sets whose union equals the universe.

In our problem, the entire sensing task set, each worker w_j's interested tasks $\mathcal{T}_j$, and its parameter μ_j corresponds to the universe, set, and weight, respectively. We aim to find the smallest weighted worker set whose interested tasks equal the entire sensing tasks. As proved in [62], the *weighted set cover problem* is NP-hard; so does the degenerated problem. Therefore, the general version of our reformulated GLMP is at least NP-hard. □

5.4 Heuristic Algorithm Design

Since it is computationally inefficient to solve the reformulated GLMP due to its NP-hardness, we propose a heuristic algorithm to derive a solution to $\boldsymbol{x}$ and $\boldsymbol{p}$. Note that they should also meet the requirement of ξ_j-*privacy*, *truthfulness* and *individual rationality*. It is also noteworthy that the removal of GLMP's probabilistic constraint (6) facilities our heuristic algorithm designs, while this can be much more complicated by solving GLMP directly.

Our heuristic algorithm mainly consists 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: $\mathcal{T}$, $\mathcal{W}$, β, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}$
Output: $\boldsymbol{x}$

$\mathcal{W}' \leftarrow \{w_j \in \mathcal{W} : \frac{b_j}{|\mathcal{T}_j|} \leq \frac{B}{|\mathcal{T}|}\}$, $\mathcal{W}_0 \leftarrow \emptyset$, $s \leftarrow 1$, $\boldsymbol{x} \leftarrow \boldsymbol{0}$, $\boldsymbol{p} \leftarrow \boldsymbol{0}$
$k \leftarrow \arg\max_{j:w_j \in \mathcal{W}' \setminus \mathcal{W}_0} \frac{g_{j|\mathcal{W}_0}}{b_j}$, $\mathcal{W}_1 \leftarrow w_k$
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**
  $x_k = 1$
  $\mathcal{W}_s \leftarrow \mathcal{W}_s \cup w_k$, $\mathcal{W}^* \leftarrow \mathcal{W}_s$, $\mathcal{T} \leftarrow \mathcal{T} \setminus \mathcal{T}_k$
  $s \leftarrow s + 1$
  $S(\mathcal{W}_s) = S(\mathcal{W}_{s-1}) + g_{s|\mathcal{W}_{s-1}}$
  $k = \arg\max_{j:w_j \in \mathcal{W}' \setminus \mathcal{W}_{s-1}} \frac{g_{j|\mathcal{W}_{s-1}}}{b_j}$
end while
Return $\boldsymbol{x}$

Winner Selection. As shown in Algorithm 1, once receiving bid profiles from all workers, the platform first rules out the workers 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 w_j's *marginal contribution* in (10) to evaluate its fitness of being selected.

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

$\mathcal{W}_{j-1}$ 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 8). 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 $S(\mathcal{W}_0) = 0$. $S(\mathcal{W}_j)$ is the sum of workers' *marginal contributions*, i.e., $g_{j|\mathcal{W}_{j-1}}$, in set $\mathcal{W}_j$. The *marginal contributions* describe how much geo-information loss is incurred for accomplishing one sensing task if the worker w_j is selected. Equation (11) is a carefully designed bid criteria to ensure that selected winning workers' final payments will not exceed platform's budget. Similar techniques are been adopted in prior works to achieve *budget feasibility* [16], [17], [21], [22]. The iteration continues until all tasks are assigned.

Payment Determination. With the selected winning workers, we follow the idea of *critical payment* [58] to determine the appropriate payments so as to achieve the requirement of *truthfulness* and *individual rationality*. 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 $\overline{\mathcal{W}}^*$.

A winner's payment is designed in a way such that it is independent with this winner's bid. It eliminates worker's incentive to manipulate bids. As shown in Theorem 7, it plays a critical role for guaranteeing truthfulness.

Algorithm Summary. Up to now, we have introduced the design of ULPT. To sum up, each worker first determines its location privacy leakage budget, based on which it generates its obfuscated location set and its mapping probability. The mean and variance of its drift distance together with its bidding profile are then sent to the platform. Note that these parameters are independent of one's true location. With the collected information from all participating workers, the platform derives the winners and their payments by solving the reformulated GLMP via our heuristic algorithm. Finally, winning workers execute their claimed tasks and send back sensing reports with their obfuscated locations embedded.

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

Proof. Since our heuristic algorithm is composed of two procedures, *winner selection* and *payment determination*, we analyze the computation complexity of these two procedures separately as follows.

In *winner selection*, the computation complexity is dominated by the while-loop of Algorithm 1 which contains $\min\{M, N\}$ iterations at most, where M and N are the number of workers and tasks, respectively. As we assume that there are much more workers than tasks in an MCS scenario, then $\min\{M, N\} = N$. For each iteration, it involves a computation for identifying the worker who has the highest $g_{j|\mathcal{W}_{s-1}}/b_j$. This computation results in M times of comparison at most. Therefore, the computation complexity for winner selection is upper bounded by $\mathcal{O}(MN)$.

In *payment determination*, each winner's payment is calculated by running the winner selection algorithm again with an updated input tuple. Thus, its computation complexity is upper bounded by $\mathcal{O}(M^2N)$.

Therefore, we conclude that the computation complexity of our heuristic algorithm is upper bounded by $\mathcal{O}(M^2N)$. □

We further derive the optimality gap caused by our heuristic algorithm.

Theorem 4. *Denote OPT as the optimal result of the reformulated GLMP, then*

$$\frac{\alpha}{OPT} \leq \frac{B \cdot \sum_{w_j \in \mathcal{W}^*} \frac{|\mathcal{T}_j|}{c_j}}{2S(\mathcal{W}^*) \min_{j\in[1,M]}\{F_j\}}, \tag{13}$$

where α is the result obtained via the heuristic algorithm.

Proof. First of all, we have $OPT \geq \min_{j\in[1,M]}\{\sigma_j/\sqrt{1-\beta} + \mu_j\} = \min_{j\in[1,M]}\{F_j\}$. Besides, via the proposed heuristic algorithm, we obtain a set of winning workers $\mathcal{W}^*$. Then α is expressed as $\alpha = \sqrt{\frac{\sum_{j:w_j\in\mathcal{W}^*}\sigma_j^2}{(1-\beta)}} + \sum_{j:w_j\in\mathcal{W}^*}\mu_j \leq \sum_{w_j\in\mathcal{W}^*} F_j = \sum_{w_j\in\mathcal{W}^*} \frac{(G(\mathcal{W}\cup w_j) - G(\mathcal{W}))}{g_{j|\mathcal{W}_{j-1}}} \leq \sum_{w_j\in\mathcal{W}^*}\frac{|\mathcal{T}_j|}{g_{j|\mathcal{W}_{j-1}}} \leq \sum_{w_j\in\mathcal{W}^*}\frac{B|\mathcal{T}_j|}{2b_jS(\mathcal{W}^*)} = \frac{B}{2S(\mathcal{W}^*)}\sum_{w_j\in\mathcal{W}^*}\frac{|\mathcal{T}_j|}{c_j}$. 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. □

6 Property Analysis

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

6.1 Privacy Protection

One of our design objectives is to guarantee ξ_j-*privacy* for each worker w_j. Recall that ξ_j is the location privacy leakage budget controlled by 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. 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

$$\begin{aligned}&\frac{\Pr[z_j|l_j]}{\Pr[z_j|l'_j]}\\ =&\frac{\exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]}{\exp[\frac{\xi_j}{r_j}(r_j - d(l'_j, z_j))]} \times \frac{\sum_{z'_j \in \mathcal{Z}'_j}\exp[\frac{\xi_j}{r_j}(r_j - d(l'_j, z'_j))]}{\sum_{\tilde{z}_j \in \mathcal{Z}_j}\exp[\frac{\xi_j}{r_j}(r_j - d(l_j, \tilde{z}_j))]}\\ =&\exp[\frac{\xi_j}{r_j}(d(l'_j, z_j) - d(l_j, z_j))] \leq \exp(\frac{\xi_j}{r_j}r_j) = \exp(\xi_j)\end{aligned}$$

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. □

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 ULPT, i.e., $\mathcal{A}_{\mathsf{Adv}}[l_j] \leq \exp[\xi_j]$.*

Proof. 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. □

Theorem 5 indicates that each winning worker has full control of its location privacy leakage to the platform during task sensing. When it has a higher 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.

6.2 Budget Feasibility

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. 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 $\overline{\mathcal{W}}^*$, 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

$$\frac{S(\mathbb{W}_2) - S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} > \frac{g_{r'|\mathbb{W}_1}}{b_{r'}}, \text{ then } \frac{S(\mathbb{W}_2) - S(\mathbb{W}_1)}{\sum_{w_t \in \mathbb{W}_2 \setminus \mathbb{W}_1} b_t} > \frac{g_{t|\mathbb{W}_1}}{b_t}$$

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),

$$\frac{g_{1|\mathcal{W}_0}}{b_1} \geq \cdots \geq \frac{g_{j|\mathcal{W}_{j-1}}}{b_j} \geq \cdots \geq \frac{g_{|\mathcal{W}^*||\mathcal{W}_{|\mathcal{W}^*|-1}}}{b_{|\mathcal{W}^*|}} \geq \frac{2S(\mathcal{W}^*)}{B}.$$

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 $p_j \leq B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)}$. □

With Lemma 2, it is straightforward to derive the following theorem.

Theorem 6. *The platform is budget feasible.*

Proof.

$$\sum_{w_j \in \mathcal{W}^*} p_j \leq \sum_{w_j \in \mathcal{W}^*} B\frac{g_{j|\mathcal{W}_{j-1}}}{S(\mathcal{W}^*)} = B,$$

i.e., the total payment to winning workers is confined to the platform's budget B. □

6.3 Economic Properties

In the following, we show that the critical economic properties, including *truthfulness* and *individual rationality*, are also achieved in ULPT.

Theorem 7. *The MCS market is truthful.*

Proof. 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 5, we derive the conclusion. □

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

Proof. 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,

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

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

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

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 6, we derive the conclusion. □

7 Relations Among Privacy, Accuracy and Budget Feasibility

From the discussion in the previous section, our mechanism guarantees ξ_j-*privacy*, (α, β)-*accuracy*, and *budget feasibility*, in addition to critical economic properties. In this section, we further explore intrinsic tradeoff relations among these three and provide insights for designing desirable MCS markets.

7.1 Privacy-Accuracy Tradeoff

Since it is difficult to derive an explicit expression for accuracy-privacy tradeoff, we leverage a term, called *privacy index*.

Definition 6. *The privacy index of an MCS market is defined as* $I = |\mathcal{H}|$, *where* $\mathcal{H} = \{w_j : \xi_j \leq \xi_0, w_j \in \mathcal{W}^*\}$, *with* ξ_0 *a given non-negative real value.*

Privacy index is in fact the number of winning workers who have relatively high privacy requirement (i.e., low privacy budget). Generally, the larger *privacy index* is, the higher privacy is guaranteed in the MCS system. Thus, it reflects the overall privacy achieved in MCS. We then have the following lemma.

Lemma 3. *Given a winning worker set $\mathcal{W}^*$, denote by $\mathcal{L}_{\mathcal{W}^*}$ and $\mathcal{L}^{(K)}_{\mathcal{W}^*}$ their two true location sets at Hamming distance $K \in [0, |\mathcal{W}^*|]$, that differs exactly on K workers' locations. Then,*

$$\Pr[loss(\mathcal{L}_{\mathcal{W}^*}) > \alpha] = \Pr[loss(\mathcal{L}^{(K)}_{\mathcal{W}^*}) > \alpha]$$

where $loss(\mathcal{L}_{\mathcal{W}^*})$ and $loss(\mathcal{L}^{(K)}_{\mathcal{W}^*})$ are geo-information loss caused by these workers when they locate at $\mathcal{L}_{\mathcal{W}^*}$ and $\mathcal{L}^{(K)}_{\mathcal{W}^*}$, respectively.

Proof. Consider $\mathcal{L}^{(k-1)}_{\mathcal{W}^*}$ and $\mathcal{L}^{(k)}_{\mathcal{W}^*}$ two location sets of $\mathcal{W}^*$ at Hamming distance 1. Let w_k be the one who locates differently, say l_k and l'_k, in these two sets. Assume that l_k and l'_k are mapped to $z_k \in \mathcal{Z}_k$ and $z'_k \in \mathcal{Z}'_k$, respectively, via w_k's location obfuscation mechanism, with $d(l_k, z_k) = d(l'_k, z'_k)$. Then,

$$\begin{aligned}\Pr[z_k|l_k] &= \frac{\exp[\frac{\xi_k}{r_k}(r_k - d(l_k, z_k))]}{\sum_{\tilde{z}_k \in \mathcal{Z}_k} \exp[\frac{\xi_k}{r_k}(r_k - d(l_k, \tilde{z}_k))]} \\ &= \frac{\exp[\frac{\xi_k}{r_k}(r_k - d(l'_k, z'_k))]}{\sum_{\hat{z}_k \in \mathcal{Z}'_k} \exp[\frac{\xi_k}{r_k}(r_k - d(l'_k, \hat{z}_k))]} = Pr[z'_k|l'_k]. \end{aligned} \quad (18)$$

Recall that the correctness of $\sum_{\tilde{z}_k \in \mathcal{Z}_k} \exp[\frac{\xi_k}{r_k}(r_k - d(l_k, \tilde{z}_k))] = \sum_{\hat{z}_k \in \mathcal{Z}'_k} \exp[\frac{\xi_k}{r_k}(r_k - d(l'_k, \hat{z}_k))]$ has been discussed in the proof for Lemma 1. Thus, the second equality above holds. (18) implies that the probability for $\mathcal{L}^{(k-1)}_{\mathcal{W}^*}$ and $\mathcal{L}^{(k)}_{\mathcal{W}^*}$ producing the same drift distance set is the same. Due to the fact that the same drift distance set leads to the same *geo-information loss* as indicated by $loss = \sum_{j:w_j \in \mathcal{W}^*} d(l_j, z_j)$, we have $\Pr[loss(\mathcal{L}^{(k-1)}_{\mathcal{W}^*}) > \alpha] = \Pr[loss(\mathcal{L}^{(k)}_{\mathcal{W}^*}) > \alpha]$. Finally, $\frac{\Pr[loss(\mathcal{L}_{\mathcal{W}^*})>\alpha]}{\Pr[loss(\mathcal{L}^{(K)}_{\mathcal{W}^*})>\alpha]} = \prod_{k=1}^{K} \frac{\Pr[loss(\mathcal{L}^{(k-1)}_{\mathcal{W}^*})>\alpha]}{\Pr[loss(\mathcal{L}^{(k)}_{\mathcal{W}^*})>\alpha]} = 1$, which ends the proof. □

Lemma 3 reveals an interesting property of our mechanism. Given the same set of winning workers, as long as their location obfuscation mechanisms stay unchanged, the probability these workers cause the same *geo-information loss* is equal even in multiple MCS auction instances where they locate differently. This is because that geo-information loss is caused by workers' location obfuscation mechanism. Though workers have different true locations, as long as they employ the same location obfuscation mechanism, their generated obfuscation locations in the set $\mathcal{Z}_j$ will have the same drifted distances and thus contribute identical geo-information loss. Therefore, two worker sets that differ exactly K locations about their genuine locations also have the same geo-information loss. Based on Lemma 3, we are ready to introduce the theorem in measuring privacy-accuracy tradeoff.

Theorem 9. ***Privacy vs. Accuracy***. *Assume that winning workers $\mathcal{W}^*$ provide $(\alpha, 1/2)$-accurate MCS services. For a given I_0, if $I \le I_0$, then $\alpha \ge \frac{r_{\max}}{4} I_0$, where $r_{\max} = \max_{j:w_j \in \mathcal{W}^*} r_j$.*

Proof. The proof is conducted via the contradiction method. Assume that $\alpha < I_0 \frac{r_{max}}{4}$. The remaining task is to elaborate that this assumption cannot exist.

Consider a winning worker set $\mathcal{H} \subset \mathcal{W}^*$ which is derived following Definition 6. We partition $\mathcal{H}$ into two parts, $\mathcal{H}_+$ and $\mathcal{H}_-$. Specifically, let $d_j = d(l_j, z_j)$ and then

$$\mathcal{H}_+ = \left\{w_j : d_j \ge \frac{r_j}{2}\right\},\ \mathcal{H}_- = \left\{w_j : d_j < \frac{r_j}{2}\right\}$$

Thus, $\mathcal{H} = \mathcal{H}_+ \cup \mathcal{H}_-$ and $\mathcal{H}_+ \cap \mathcal{H}_- = \emptyset$. There must be one set, either $\mathcal{H}_+$ or $\mathcal{H}_-$, such that its privacy index is at most $I/2$. Without loss of generality, we assume $I(\mathcal{H}_+) \le I/2$. Consider $\mathcal{H}$'s two location sets $\mathcal{L}_{\mathcal{H}}$ and $\mathcal{L}'_{\mathcal{H}}$, where $\mathcal{L}'_{\mathcal{H}}$ is formulated in the following way.

If $w_j \in \mathcal{H}_-$, then $d'_j = d_j$. If $w_j \in \mathcal{H}_+$, then

$$d'_j = \begin{cases} d_j - \frac{r_j}{2}, & \text{if } d_j \ge \frac{r_j}{2} \\ d_j + \frac{r_j}{2}, & \text{if } d_j < \frac{r_j}{2}. \end{cases}$$

We have

$$\begin{aligned} |loss(\mathcal{L}_{\mathcal{H}}) - loss(\mathcal{L}'_{\mathcal{H}})| &= \left| \sum_{j:w_j \in \mathcal{H}_+} (d_j - d'_j) \right| \\ = \frac{1}{2} \left| \sum_{j:w_j \in \mathcal{H}_+} r_j \right| &\le \frac{r_{\max}}{2} I(\mathcal{H}_+) \le \frac{r_{\max}}{4} I \le \frac{r_{\max}}{4} I_0. \end{aligned} \quad (19)$$

Since $\mathcal{L}_{\mathcal{H}}$ and $\mathcal{L}'_{\mathcal{H}}$ have the Hamming distance of $|\mathcal{H}_+|$, by implementing Lemma 3 and the statement that $\mathcal{W}^*$ provides $(\alpha, 1/2)$-accurate services, then

$$\begin{aligned} \Pr[loss(\mathcal{L}'_{\mathcal{H}}) > \alpha] &= \Pr[loss(\mathcal{L}_{\mathcal{H}}) > \alpha] \\ &\le \Pr[loss(\mathcal{L}_{\mathcal{W}^*}) > \alpha] < \frac{1}{2}. \end{aligned}$$

Together with $loss(\mathcal{L}_{\mathcal{H}}) \ge loss(\mathcal{L}'_{\mathcal{H}}) - \frac{r_{\max}}{4} I_0$ indicated by (19), then $\Pr[loss(\mathcal{L}_{\mathcal{H}}) > \alpha - \frac{r_{\max}}{4} I_0] \le 1/2$. According to the discussion above,

$$\Pr[\alpha - \frac{r_{\max}}{4} I_0 < loss(\mathcal{L}_{\mathcal{H}}) \le \alpha] < 1/2. \quad (20)$$

On the other hand, since $\alpha < \frac{r_{\max}}{4} I_0$, then $\Pr[loss(\mathcal{L}_{\mathcal{H}}) > \alpha - \frac{r_{\max}}{4} I_0] = 1$. Besides, $\Pr[loss(\mathcal{L}_{\mathcal{H}}) \le \alpha] \ge \Pr[loss(\mathcal{L}_{\mathcal{W}^*}) \le \alpha] \ge 1/2$. Thus

$$\Pr[\alpha - I_0 \frac{r_{\max}}{4} < loss(\mathcal{L}_{\mathcal{H}}) \le \alpha] \ge 1/2. \quad (21)$$

Apparently, (20) and (21) contradict with each other. Therefore, the assumption that $\alpha < \frac{r_{\max}}{4} I_0$ is wrong, and thus $\alpha \ge \frac{r_{\max}}{4} I_0$. □

Theorem 9 describes the relation between the overall privacy achieved by winning workers and service accuracy. Generally, when better overall privacy is achieved, the lower service accuracy the platform provides.

7.2 Privacy-Budget Tradeoff

Theorem 10. ***Privacy vs. Budget Feasibility***. *For a given budget B, the privacy leakage budget from each winning worker $w_j \in \mathcal{W}^*$ should satisfy*

$$\xi_j \le (B/2 - c^s_j)/c^p_j. \quad (22)$$

Proof. This statement can be easily inferred from Theorem 7 and (16). Specifically, $c_j = c^s_j + c^p_j \xi_j = b_j \le B/2$, and thus $\xi_j \le (B/2 - c^s_j)/c^p_j$. □

From the perspective of the platform, when it has a small budget amount, it can only afford to select workers who impose high privacy requirement, i.e., low privacy leakage budget ξ_j, to execute sensing tasks. This is intuitive that

these workers upload sensing reports with more noisy locations and thus ask for less compensation for their privacy loss.

The result from Theorem 10 can also be leveraged to accelerate an auction process. Before an auction begins, the platform broadcasts its budget B to all workers. Upon receiving this message, each worker finds out if its privacy leakage budget meets the condition specified by (22). If no, this worker is less likely to win in this round of auction. Thus it is wiser to wait until the platform increases its budget. With the decreased number of participating workers, *winner selection* and *payment determination* would be conducted more efficiently.

7.3 Accuracy-Budget Tradeoff

Theorem 11. ***Quality vs. Budget Feasibility****. For a given budget B and confidence level β, the platform provides (α, β)-accurate services with*

$$\alpha \geq \sum_{w_j \in \mathcal{W}^*} r_j \left(1 + \frac{1}{2}(1 - \sqrt[|\mathcal{W}^*|]{\beta}) \exp[\frac{B/2 - c_j^s}{2c_j^p}]\right)^{-1}.$$

Proof. For an arbitrary winning worker $w_j \in \mathcal{W}^*$ who locates at l_j, we define $S_{d_j/2} = \{z_j : d(l_j, z_j) < d_j/2\}$ and $S_{d_j} = \{z_j : d(l_j, z_j) > d_j\}$, where d_j is a parameter taking value from $[\Delta r_j, r_j]$.

$$\begin{aligned}
&\frac{\Pr[S_{d_j}]}{\Pr[S_{d_j/2}]} \\
=&\frac{\sum\limits_{z_j \in S_{d_j}} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]}{\sum\limits_{z'_j \in \mathcal{Z}_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z'_j))]} \cdot \frac{\sum\limits_{z'_j \in \mathcal{Z}_j} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z'_j))]}{\sum\limits_{z_j \in S_{d_j/2}} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]} \\
=&\frac{\sum\limits_{z_j \in S_{d_j}} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]}{\sum\limits_{z_j \in S_{d_j/2}} \exp[\frac{\xi_j}{r_j}(r_j - d(l_j, z_j))]} \overset{①}{\leq} \frac{|S_{d_j}| \exp[\frac{\xi_j}{r_j}(r_j - d_j)]}{|S_{d_j/2}| \exp[\frac{\xi_j}{r_j}(r_j - \frac{d_j}{2})]} \\
=&\frac{|S_{d_j}|}{|S_{d_j/2}|} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}] \overset{②}{\leq} \frac{r_j - d_j}{\Delta r_j} \times \frac{2\Delta r_j}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}] \\
=&\frac{2(r_j - d_j)}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}]
\end{aligned}$$

where ① comes from the definition of $S_{d_j/2}$ and S_{d_j}. ② is derived from the generation of $\mathcal{Z}_j$; there are $\frac{r_j - d_j}{\Delta r_j} \cdot \frac{2\pi}{\Delta\theta_j}$ and $\frac{d_j/2}{\Delta r_j} \cdot \frac{2\pi}{\Delta\theta_j}$ elements in S_{d_j} and $S_{d_j/2}$, respectively.

Since $\Pr[S_{d_j}] \leq \Pr[S_{d_j}]/\Pr[S_{d_j/2}]$, then $\Pr[S_{d_j}] \leq \frac{2(r_j - d_j)}{d_j} \exp(-\frac{\xi_j}{r_j}\frac{d_j}{2})$, and thus $1 - \Pr[S_{d_j}] \geq 1 - \frac{2(r_j - d_j)}{d_j} \exp(-\frac{\xi_j}{r_j}\frac{d_j}{2})$, i.e.,

$$\Pr[d(l_j, z_j) \leq d_j] \geq 1 - \frac{2(r_j - d_j)}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}].$$

Now, let $1 - \sqrt[|\mathcal{W}^*|]{\beta} = \frac{2(r_j - d_j)}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}]$. Then

$$\prod_{w_j \in \mathcal{W}^*} \Pr[d(l_j, z_j) \leq d_j] \geq \beta,$$

which implies

$$\Pr[\sum_{w_j \in \mathcal{W}^*} d(l_j, z_j) \leq \sum_{w_j \in \mathcal{W}^*} d_j] = \Pr[loss \leq \sum_{w_j \in \mathcal{W}^*} d_j] \geq \beta.$$

Let $\alpha = \sum_{w_j \in \mathcal{W}^*} d_j$. According to Definition 2, the mechanism achieves (α, β) service accuracy. On the other hand,

$$\begin{aligned}
1 - \sqrt[|\mathcal{W}^*|]{\beta} &= \frac{2(r_j - d_j)}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{d_j}{2}] \\
&\geq \frac{2(r_j - d_j)}{d_j} \exp[-\frac{\xi_j}{r_j}\frac{r_j}{2}] \\
\Rightarrow d_j &\geq r_j \left(1 + \frac{1}{2}(1 - \sqrt[|\mathcal{W}^*|]{\beta}) \exp[\frac{\xi_j}{2}]\right)^{-1} \qquad (23)
\end{aligned}$$

Then, we can have the conclusion that

$$\Pr[d(l_j, z_j) \leq d_j] \geq \sqrt[|\mathcal{W}^*|]{\beta}$$

in which $d_j \geq r_j \left(1 + \frac{1}{2}(1 - \sqrt[|\mathcal{W}^*|]{\beta}) exp(\frac{\xi_j}{2})\right)^{-1}$.

Applying this intermediate conclusion to all winning workers, we can have that Based on (22) and (23),

$$\begin{aligned}
\alpha = \sum_{w_j \in \mathcal{W}^*} d_j &\geq \sum_{w_j \in \mathcal{W}^*} r_j \left(1 + \frac{1}{2}(1 - \sqrt[|\mathcal{W}^*|]{\beta}) \exp[\frac{\xi_j}{2}]\right)^{-1} \\
&\geq \sum_{w_j \in \mathcal{W}^*} r_j \left(1 + \frac{1}{2}(1 - \sqrt[|\mathcal{W}^*|]{\beta}) \exp[\frac{B/2 - c_j^s}{2c_j^p}]\right)^{-1}
\end{aligned}$$

The proof ends. □

Theorem 11 tells that α has a negative correlation with budget B. Recall that α is the upper bound of *geo-information loss* provided by the platform for a given β. Therefore, in order to achieve a lower α, thus potentially a smaller *geo-information loss*, the platform needs to prepare for a larger amount of budget B. This statement meets our expectation–better service accuracy requires the platform to recruit workers with larger privacy leakage budgets which, in turn, ask for higher privacy loss compensation.

8 Performance Evaluation

In this section, we conduct a series of simulations to evaluate performances of the proposed ULPT. Follow prior works on MCS [18]–[20], we employ the New York City's 311 dataset [59], which contains non-emergency reports of the citizens. It allows 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 focus on noise complaints. According to 311 records, noise is the third largest category of complaints. As shown in Fig. 3, the circles represent noise complains received in that area. A darker a circle indicates more complains observed. When complaining about noises, people are required to provide the location, time, and a fine-grained noise category, such as loud music or construction. Hence, the 311 complaint data about noises can be viewed as a result of crowd sensing, leveraging "human as a sensor". As a note, 311 dataset has been widely adopted in social/crowd sensing related research.

In the evaluation, 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. We further set the rest

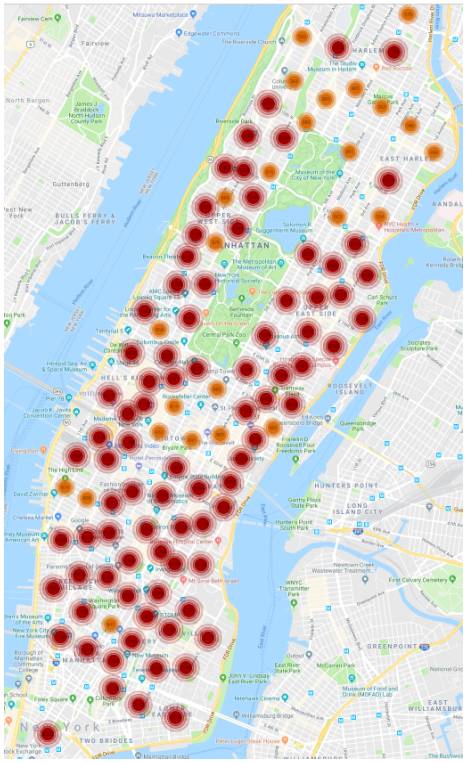
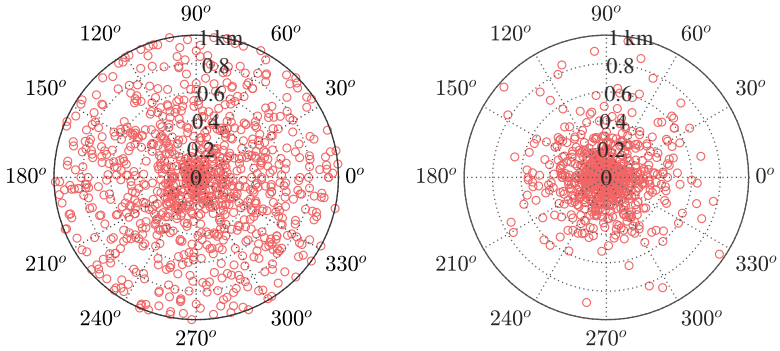
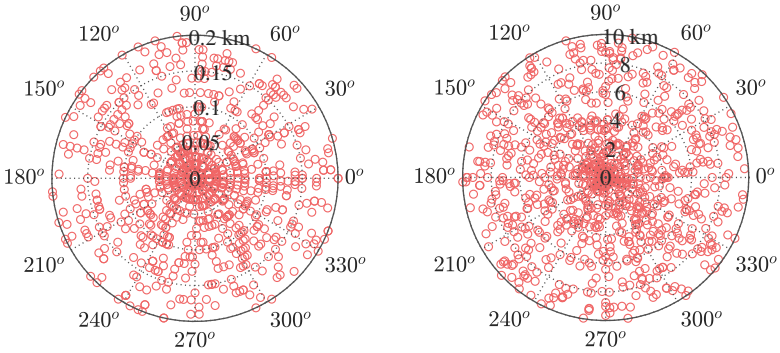


Fig. 3: The noise complain distribution in the Manhattan area according to New York City's 311 database.



(a) Privacy budget $\xi_j = 0.1$. (b) Privacy budget $\xi_j = 5$.

Fig. 4: Obfuscated locations under different privacy budgets.



(a) Concerned radius $r_j = 0.2$ km. (b) Concerned radius $r_j = 10$ km.

Fig. 5: Obfuscated locations under different location obfuscation ranges.

parameters as follows. The numbers in bold are the default values if not mentioned otherwise. A worker's obfuscation range r_j ranges between $[1, \cdots, \mathbf{5}, \cdots, 10]$km. Its location privacy leakage budget ξ_j is selected from $[0.01, 0.02, \cdots, \mathbf{0.1}]$. Besides, the platform's budget B takes value from $[20, \cdots, \mathbf{100}]$. The confidence level is set to $\beta = [0.3, 0.5, 0.9, \mathbf{0.95}]$. Our code runs on a desktop with 3.4GHz Intel i7 CPU and 16 GB memory. All simulation results are the average over 100 trials.

8.1 Performance Benchmarks

8.1.1 Performances of Our Location Obfuscation Mechanism

The proposed location obfuscation mechanism generates an alternative location to substitute worker w_j's genuine one so as to protect its location privacy from the platform. As discussed in section 5.2, workers are allowed to customize its own obfuscation mechanism by selecting different obfuscation radius r_j and privacy budgets ξ_j. We then evaluate the impact of these parameters to the geographical distributions of the obfuscated locations.

Impact of Privacy Leakage Budget ξ_j. We consider both scenarios of a randomly selected worker that has significantly different privacy budgets, i.e., $\xi_j = 0.1$ and $\xi_j = 5$, respectively. Fig. 4 depicts w_j's obfuscated locations generated in its established polar coordinate system with a fixed $r_j = 1km$ in 1000 runs. We observe the obfuscated locations z_j are more concentrated around the w_j's genuine location (which is the center of the polar coordinates), when a larger privacy budget is chosen. This is because the probabilistic mapping function (7) tends to choose the locations that are close to l_j. It would be even more biased as the increase of ξ_j. Clearly, location privacy is better protected under a smaller ξ_j.

Impact of Obfuscation Range r_j. Fig. 5 further depicts the impact of obfuscation range r_j to the geographical distributions of the obfuscated locations given $\epsilon_j = 0.1$. We observe that the locations generated in Fig. 5(b) is more evenly distributed than those of Fig. 5(a). This is because privacy budget $\xi_j = \epsilon_j / r_j$ is much smaller by selecting a larger r_j. Following (7), the chances of choosing different locations within the location obfuscation set are more likely to be equal under small privacy budgets ξ_j.

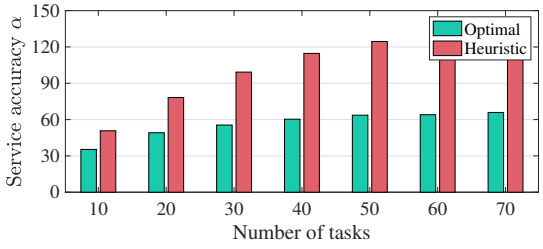
8.1.2 Performances of Our Heuristic Algorithm

We then 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. We enumerate each possible solution of the reformulated GLMP and identify the one producing the minimum value α. Note that CPLEX optimizer [61], the conventional tool to solve standard optimization problems, is inapplicable here due to the square root involved in the objective function.

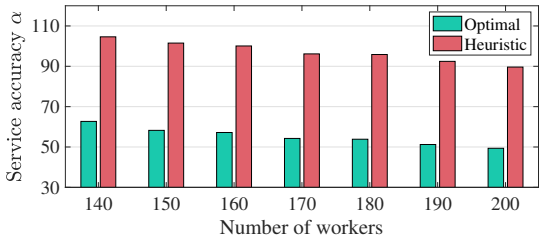
Optimality Gap. The impact of task size N and worker size M are examined in Fig. 6(a) and Fig. 6(b), respectively. We observe in Fig. 6(a) that α increases as the number of task grows. This is because when there are more tasks, more winning workers are recruited, which leads to larger

TABLE 2: Comparison between exhaustive search and the heuristic algorithm in terms of computation efficiency.

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

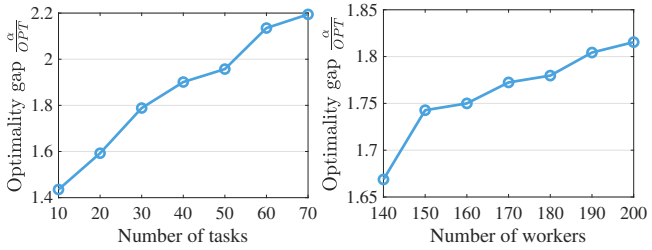


(a) Impact of task size N



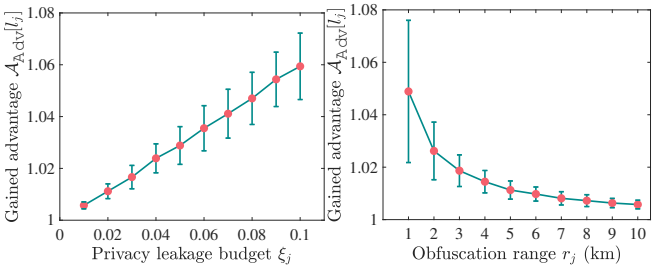
(b) Impact of worker size M

Fig. 6: Optimality gap of the heuristic algorithm under different settings.



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

Fig. 7: Optimality gap ratio under different settings.



(a) Impact of ξ_j. (b) Impact of r_j.

Fig. 8: Platform's posterior knowledge gain with respect to worker's privacy leakage budget and obfuscation range.

accumulated *geo-information loss*. Meanwhile, we have an opposite observation with respect to the number of workers in Fig. 6(b). This is because the platform can potentially choose from a larger audience of workers with larger privacy leakage budgets that cause less *geo-information loss*. In addition, we note that the heuristic algorithm generally have higher α than the optimal one. For example, when $N = 20$, $\alpha = 73.46$ for the former, while $\alpha = 48.75$ for the latter. This optimization gap is a result of two reasons: Firstly, in trade of shorter computation time, the heuristic algorithm gives the computation accuracy away to search for a good solution of the GLMP which is not necessarily the optimal one. 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.

Fig. 7 further shows the optimality gap with respect to the task size N and worker size M. It is observed that the gap increases with the growth of N and M. Hence, it is harder for the heuristic algorithm to derive a close-to-optimal solution in large MCS markets. For example, the gap is 1.43 when the task size $N = 10$; it increases to 1.9 when $N = 40$. This is because the upper bound of the optimality gap is positively correlated with the task size N as indicated by equation (13).

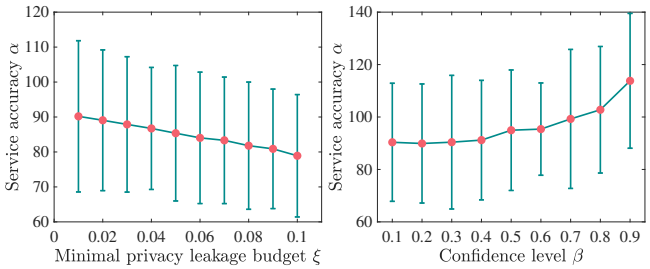
Computation Efficiency. Table 1 compares the computation time for both algorithms under different MCS market sizes. Particularly, under the setting $M = 190,\ N = 100$, it only costs 309.06 ms for the heuristic algorithm to find the solution, while that for the exhaustive search is significantly larger,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

8.2 Privacy Protection

To evaluate the performance of location privacy preserving, we examine the platform's posterior knowledge gain $\mathcal{A}_{\mathsf{Adv}}[l_j]$ towards workers' genuine location l_j given the reported location z_j. An arbitrary worker w_j is randomly selected and is allowed to determine its own location obfuscation mechanism, which includes selecting the privacy leakage budget ξ_j and obfuscation range r_j. We then evaluate the impact of these two parameters to $\mathcal{A}_{\mathsf{Adv}}[l_j]$.

Impact of Privacy Leakage Budget ξ_j. Fig. 8(a) depicts the platform's posterior knowledge gain $\mathcal{A}_{\mathsf{Adv}}[l_j]$ when w_j chooses different ξ_j with a fixed $r_j = 10$km. We observe that $\mathcal{A}_{\mathsf{Adv}}[l_j]$ 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 $\xi_j = 0.07$. It validates the theoretical results 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 less advantage to correctly infer the worker's location l_j.

Impact of Obfuscation Range r_j. Fig. 8(b) further depicts the impact of obfuscation range r_j to $\mathcal{A}_{\mathsf{Adv}}[l_j]$ under a fixed $\epsilon_j = 0.1$. It is observed that with the increase of worker's obfuscation range r_j, the $\mathcal{A}_{\mathsf{Adv}}[l_j]$ decreases. This is because $\xi_j = \epsilon_j / r_j$ is smaller under the larger r_j. And, following the Theorem 5, the worker's location can be better preserved when a small ξ_j is selected.



(a) Impact of ξ. (b) Impact of β.

Fig. 9: Platform's MCS service accuracy with respect to β and workers' privacy leakage budgets.

8.3 MCS Service Accuracy

We evaluate the service accuracy by examining α with respect to ξ and β, respectively. 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. In the experiments, we intentionally use ξ instead of ξ_j in the figure. In specific, as ξ_j decreases, it means w_j tends to add more noise to its location. According to our design, it becomes less likely to be selected as a winning worker. In this case, it is pointless to consider the relation between the associated ξ_j with α. To resolve this issue, we employ ξ to reflect the privacy leakage budgets ξ_j's from individual winning workers instead; when they have smaller budgets, ξ becomes small accordingly, and vise versa.

Fig. 9(a) implies 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 $\xi = 0.09$. It coincides with the conclusion of Theorem 9. When ξ_j increases, i.e., every worker has a larger privacy leakage budget, ξ increases as well, which leads to the decrease of the privacy index I of an MCS market. Besides, for a given I_0, if $I \leq I_0$, then $\alpha \geq \frac{r_{\max}}{4} I_0$. Therefore, a larger ξ results in a smaller α.

Impact of β. Fig. 9(b) depicts the relation between α and β. We observe that α increases as the growth of β. For example, $\alpha = 0.23$ when $\beta = 0.3$ and $\alpha = 0.34$ when $\beta = 0.9$. Recall that β is the confidence level for the statement $loss \leq \alpha$. Then, when we want to estimate service accuracy with higher confidence, a more conservative and thus a relatively large α will be given.

8.4 Budget Feasibility

In this section, we evaluate the impact of platform's budget to MCS market performances.

Impact to $\mathcal{A}_{\mathsf{Adv}}[l_j]$. Fig. 10(a) examines the platform's advantage in inferring two arbitrary workers' exact locations, whose privacy budget is set as 0.01 and 0.05, respectively. We observe that the $\mathcal{A}_{\mathsf{Adv}}[l_j]$ is independent from the budget B. This is because each worker independently determines their location privacy budget ξ_j, which implies how much advantages the platform can gain about the true location l_j. This value is not correlated to the platform's budget.

Impact to service accuracy α. Fig. 10(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.

Impact to Privacy Index I. Fig. 10(c) clearly illustrates a tradeoff relation between B and the market overall privacy, represented by privacy index I. I decreases as the increase of B. This is because the platform can hire workers who have larger privacy leakage budgets with a larger B. Consequently, the overall privacy achieved by the market becomes loose. Meanwhile, the overall privacy experiences stable changes when B is larger than 60. Therefore, if the platform also emphasizes over worker's location privacy, it should confines its budget within 60 in an auction.

Impact to Winning Workers' Payments. Fig. 10(d) shows the impact of platform's budget to winning workers' payments. Specifically, we examine three winning workers with different privacy leakage budget at 0.01 and 0.05. First of all, we notice that both payments increase as B grows. It meets our payment determination procedure. Besides, we also notice that the worker with a higher privacy leakage budget are paid more. For example, when $B = 40$, worker 1 is paid as 5.56 while worker 2 is paid as 8.68. This is because the worker discloses more information regarding its true location during sensing result reporting. As a result, they ask for higher payment to compensate their privacy loss. Therefore, if a worker has loose location privacy requirement, it can choose to report a less obfuscated location and ask for a higher payment from the platform.

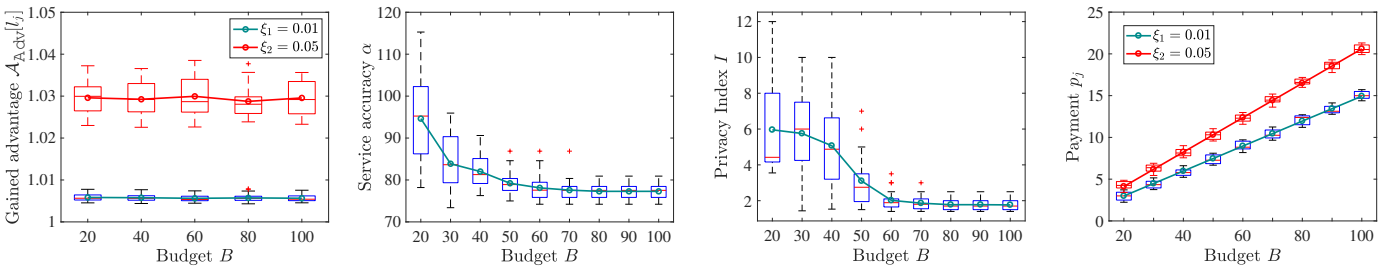
8.5 Relations Among Privacy, Accuracy and Budget

We further illustrate in Fig. 11 the relations among privacy, accuracy and budget achieved by our mechanism. Different confidence levels of β are considered, $\beta = [0.3, 0.5, 0.9]$. The 3D curve is projected to $B - \alpha$ plane and $\alpha - I$ plane. Its projection to the $B - \alpha$ plane exactly shows the tradeoff between the platform's budget B and service accuracy α. When the platform has a larger budget, the service accuracy (and thus the inverse of α) increases. The 3D curve's projection to the $\alpha - I$ plane shows the tradeoff between the service accuracy and the overall achieved privacy. Particularly, when the platform aims to provide services with better accuracy, the privacy index I diminishes, i.e., there are less winning workers with strong privacy protection.

8.6 Worker's spatial distribution

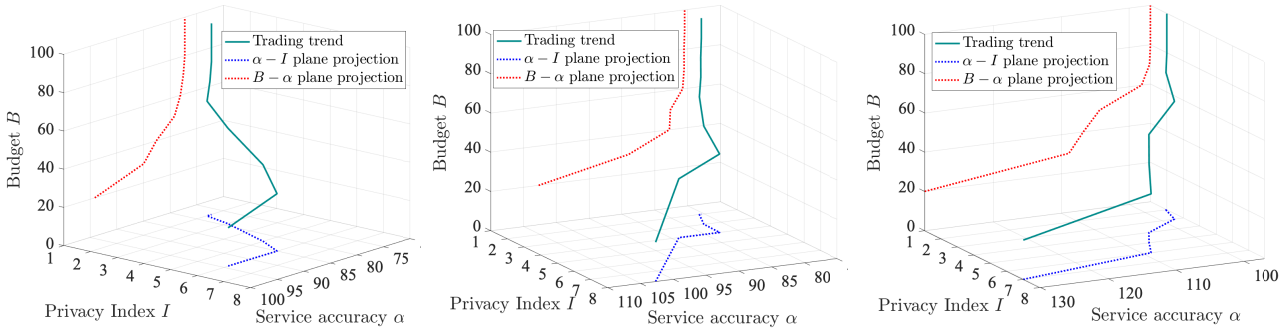
We evaluate the impact of worker's spatial distributions by settings workers' locations within different distribution ranges. A small distribution range represents workers are located within close proximity and vice versa. As shown in Fig. 12, there is no noticeable advantage for the adversary to infer worker's true locations. This is because each worker protects its location independently following its own privacy budget ξ_j. As proved in Theorem 5, this advantage is confined within $\exp[\xi_j]$. In conclusion, the changes in worker's spatial distribution do not impact the adversary's gained advantage.

Workers make their reports individually and independently. Besides, their locations are obfuscated locally by applying the proposed location obfuscation mechanism with



(a) Impact to $\mathcal{A}_{\mathsf{Adv}}(l_j)$. (b) Impact to service accuracy α. (c) Impact to privacy index I. (d) Impact to winning workers' payments.

Fig. 10: Impact of the platform's budget to MCS market performances.



(a) $\beta = 0.3$. (b) $\beta = 0.5$. (c) $\beta = 0.9$.

Fig. 11: Evaluating relations among privacy, accuracy and budget.

their customized privacy-preserving budgets ξ_j's, regardless of peer worker's locations or strategies. Each worker w_j's reported location is randomly selected from its own location obfuscation set $\mathcal{Z}_j$. Even multiple workers are within close proximity, their obfuscated locations can be significantly distinct from each other. Thus, the a combination of multiple random locations would not provide any extra advantage to the platform.

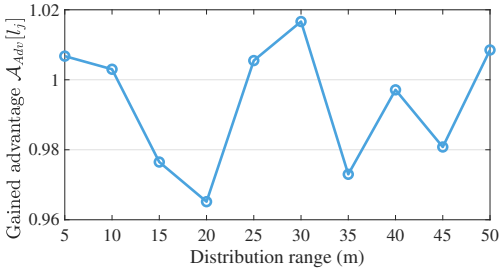


Fig. 12: Impact of worker's spatial distributions.

9 Discussions

As suggested by [23]–[27], worker's sensing data can potentially disclose critical private information such as locations, trajectories, habits, preferences, and so on. To resolve this issue, extensive prior efforts have been devoted to protecting data privacy in the context of MCS [4], [5], [47]–[53]. Unlike these works, here we focus on mitigating worker's location information leakage caused by direct embedding their geographic positions in their sensory reports. We believe existing schemes on protecting sensing data privacy can be easily adapted to our framework. Meanwhile, our paper does consider the impact of correlation between sensing data and location information. Specifically, we formalize the adversary's prior knowledge regarding a worker's genuine location into a probability distribution $\Pr[l_j]$. Part of the prior knowledge can come from the side-channel mentioned above. This work aims to ensure that the adversary cannot gain extra information on a worker's location by observing its obfuscated location in the sensing report.

One limitation of our scheme is the privacy leakage caused by continuous reporting. Our location protection scheme is developed based on the framework of differential privacy (DP). Hence, it heritages the limitation of DP in some aspects. The most relevant one is extra privacy leakage caused by continuous reporting. Specifically, from the perspective of the platform, w_j's equivalent value of ξ_j increases as the worker submits multiple perturbed locations, all generated under ξ_j-privacy. Such a property is called *sequential compositions* under differential privacy [14], [15]. As a possible approach, a worker can start from a lower privacy leakage budget $\xi'_j < \xi_j$, where ξ_j is the worker's target budget. ξ'_j can be calculated if the worker has an accurate estimation of the number of sensing tasks T it aims to attend. Then the target ξ_j will be achieved after T rounds of submission with obfuscated location z_j perturbed under ξ'_j.

10 Conclusion

We construct a location privacy trading framework, ULPT, for MCS with an auction approach, where the platform provides incentives to motivate workers to com-

plete sensing tasks. We incorporate the notation of *geo-distinguishability* in our mechanism design. Workers with different location privacy leakage budgets have freedom to choose how much privacy to disclose to the platform, in trade of different monetary reward. Taking into account of budget constraint, service accuracy, and privacy protection, we formulate an optimization problem, which is proved to be at least NP-hard. To efficiently solve it, a heuristic algorithm is proposed, with bounded computation complexity and optimality gap. More importantly, the paper provides a rigorous theoretical analysis over the design objectives, including ξ-*privacy*, (α, β)-*accuracy*, *budget feasibility*, as well as their comprehensive tradeoff relations. All these theoretical results have been validated through extensive simulations based on New York City's 311 platform dataset.

ACKNOWLEDGMENTS

We sincerely thank the anonymous reviewers for their insightful comments and suggestions. The work of M. Li is partially supported by NSF CNS-1943509 and ECCS-1849860. The work of L. Guo is partially supported by NSF under grant IIS-1949640 and CNS-2008049. The work of L. Yang is supported in part by NSF IIS-1838024, EEC-1801727, and CNS-1950485.

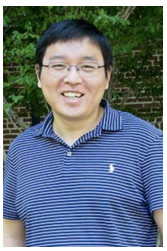
REFERENCES

[1] Myheartmap. [Online]. Available: http://www.med.upenn.edu/
[2] L. Wang, et al., "Sparse Mobile Crowdsensing with Differential and Distortion Location Privacy," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2735–2749, Feb 2020.
[3] X. Jin, et al., "Privacy-preserving crowdsourced spectrum sensing," *IEEE/ACM Transactions on Networking*, vol. 26, pp. 1236–1249, Jun 2018.
[4] G. Yang, et al., "Socially Privacy-preserving Data Collection for Crowdsensing," *IEEE Transactions on Vehicular Technology*, vol. 69, pp. 851–861, Nov 2019.
[5] G. Liao, et al., "Prospect Theoretic Analysis of Privacy-Preserving Mechanism," *IEEE Transactions on Vehicular Technology*, vol. 28, pp. 71–83, Feb 2020.
[6] C. Niu, et al., "Unlocking the value of privacy: Trading aggregate statistics over private correlated data," in *Proceedings of ACM SIGKDD*, 2018.
[7] Z. Zhang, S. He, J. Chen, and J. Zhang, "Reap: An efficient incentive mechanism for reconciling aggregation accuracy and individual privacy in crowdsensing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 2995–3007, 2018.
[8] P. Sun, Z. Wang, Y. Feng, L. Wu, Y. Li, H. Qi, and Z. Wang, "Towards personalized privacy-preserving incentive for truth discovery in crowdsourced binary-choice question answering," in *Proceedings of the IEEE Conference on Computer Communications*, 2020.
[9] J. Xiong, R. Ma, L. Chen, Y. Tian, Q. Li, X. Liu, and Z. Yao, "A personalized privacy protection framework for mobile crowdsensing in iiot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4231–4241, 2019.
[10] L. Wang, D. Zhang, D. Yang, B. Y. Lim, X. Han, and X. Ma, "Sparse mobile crowdsensing with differential and distortion location privacy," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2735–2749, 2020.
[11] J. Gao, S. Fu, Y. Luo, and T. Xie, "Location privacy-preserving truth discovery in mobile crowd sensing," in *Proceedings of the International Conference on Computer Communications and Networks*, 2020.
[12] X. Yi, K.-Y. Lam, E. Bertino, and F.-Y. Rao, "Location privacy-preserving mobile crowd sensing with anonymous reputation," in *Proceedings of the European Symposium on Research in Computer Security*, 2019.
[13] E. S. Domi and G. Usha, "Tsdla: Algorithm for location privacy in clustered lb-mcs network," in *Proceedings of the International Conference on Smart Systems and Inventive Technology*, 2020.
[14] C. Dwork, "Differential privacy: A survey of results," in *Proceedings of the International conference on theory and applications of models of computation*, 2008.
[15] C. Dwork and A. Roth, "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*, vol. 9, no. 3-4, pp. 211–407, 2014.
[16] Y. Singer, "Budget feasible mechanisms," in *Proceedings of the Symposium on Foundations of Computer Science*, 2010.
[17] N. Anari, G. Goel, and A. Nikzad, "Mechanism design for crowdsourcing: An optimal 1-1/e competitive budget-feasible mechanism for large markets," in *Proceedings of the IEEE Symposium on Foundations of Computer Science*, 2014.
[18] C. Huang, X. Wu, and D. Wang, "Crowdsourcing-based urban anomaly prediction system for smart cities," in *Proceedings of the ACM international on conference on information and knowledge management*, 2016, pp. 1969–1972.
[19] R. Solymosi, K. J. Bowers, and T. Fujiyama, "Crowdsourcing subjective perceptions of neighbourhood disorder: Interpreting bias in open data," *The British Journal of Criminology*, vol. 58, no. 4, pp. 944–967, 2018.
[20] B. Y. Clark and J. L. Brudney, "Citizen representation in city government-driven crowdsourcing," *Computer Supported Cooperative Work*, vol. 28, no. 5, pp. 883–910, 2019.
[21] B. Song, H. Shah-Mansouri, and V. W. Wong, "Quality of sensing aware budget feasible mechanism for mobile crowdsensing," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3619–3631, 2017.
[22] X. Zhang, Z. Yang, Y. Liu, J. Li, and Z. Ming, "Toward efficient mechanisms for mobile crowdsensing," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 2, pp. 1760–1771, 2016.
[23] J. Han, E. Owusu, L. T. Nguyen, A. Perrig, and J. Zhang, "Accomplice: Location inference using accelerometers on smartphones," in *Proceedings of the International Conference on Communication Systems and Networks*, 2012.
[24] S. Nawaz and C. Mascolo, "Mining users' significant driving routes with low-power sensors," in *Proceedings of the ACM Conference on Embedded Network Sensor Systems*, 2014.
[25] X. Zhou, S. Demetriou, D. He, M. Naveed, X. Pan, X. Wang, C. A. Gunter, and K. Nahrstedt, "Identity, location, disease and more: Inferring your secrets from android public resources," in *Proceedings of the ACM SIGSAC conference on Computer & communications security*, 2013.
[26] Y. Michalevsky, A. Schulman, G. A. Veerapandian, D. Boneh, and G. Nakibly, "Powerspy: Location tracking using mobile device power analysis," in *Proceedings of the {USENIX} Security Symposium*, 2015.
[27] S. Narain, T. D. Vo-Huu, K. Block, and G. Noubir, "Inferring user routes and locations using zero-permission mobile sensors," in *Proceedings of the IEEE Symposium on Security and Privacy*, 2016.
[28] "Waze." [Online]. Available: https://www.waze.com/
[29] "Noisecapture." [Online]. Available: http://noise-planet.org/noisecapture.html
[30] C. Cornelius, et al., "Anonysense: privacy-aware people-centric sensing," in *Proceedings of ACM Mobisys*, 2008.
[31] L. Pournajaf, et al., "Spatial task assignment for crowd sensing with cloaked locations," in *Proceedings of the IEEE MDM*, 2014.
[32] I. J. Vergara-Laurens, et al., "Privacy, quality of information, and energy consumption in participatory sensing systems," in *Proceedings of IEEE PerCom*.
[33] L. Wang, et al., "Differential location privacy for sparse mobile crowdsensing," in *Proceedings of IEEE ICDM*, 2016.
[34] L. Wang, et al., "Location privacy-preserving task allocation for mobile crowdsensing with differential geo-obfuscation," in *Proceedings of the WWW*, 2017.
[35] "acxiom." [Online]. Available: https://www.acxiom.com/
[36] A. Ghosh and A. Roth, "Selling privacy at auction," *Games and Economic Behavior*, vol. 91, pp. 334–346, May 2015.
[37] 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.
[38] A. Ghosh, et al., "Privacy and coordination: computing on databases with endogenous participation," in *Proceedings of ACM EC*, 2013.

[39] A. Ghosh, et al., "Buying private data without verification," in *Proceedings of ACM EC*, 2014.
[40] K. Nissim, et al., "Redrawing the boundaries on purchasing data from privacy-sensitive individuals," in *Proceedings of ACM ITCS*, 2014.
[41] W. Wang, et al., "Buying data from privacy-aware individuals: the effect of negative payments," in *Proceedings of WINE*, 2016.
[42] ——, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *Proceedings of the ACM SIGMETRICS*, 2016.
[43] M. E. Andrés, et al., "Geo-indistinguishability: Differential privacy for location-based systems," in *Proceedings of ACM CCS*, 2013.
[44] 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.
[45] X. Wang, et al., "Incentivizing crowdsensing with location-privacy preserving," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6940–6952, October 2017.
[46] Q. Li and G. Cao, "Efficient privacy-preserving stream aggregation in mobile sensing with low aggregation error," in *Proceedings on PETS*, 2013.
[47] D. Wu, et al., "Dynamic Trust Relationships Aware Data Privacy Protection in Mobile Crowd-Sensing " *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2958–2970, August 2018.
[48] E. Cristofaro and C. Soriente, "Participatory privacy: Enabling privacy in participatory sensing ," *IEEE Network*, vol. 27, no. 1, pp.32–36, February 2013.
[49] Y. Sei and A. Ohsuga, "Differential private data collection and analysis based on randomized multiple dummies for untrusted mobile crowdsensing," *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 4, pp. 926–939, April 2017.
[50] M. Groat, et al., "Enhancing privacy in participatory sensing applications with multidimensional data," in *Proceedings of the IEEE PerCom*, 2012.
[51] S. Goryczka and L. Xiong, "A comprehensive comparison of multiparty secure additions with differential privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 14, no. 5, pp. 463–477, September 2017.
[52] C. Cornelius, et al., "Anonysense: privacy-aware people-centric sensing, " in *Proceedings of the ACM MobiSys*, 2008.
[53] 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, June 2015.
[54] H. Jin, et al., "Inception: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proceedings of the ACM MobiHoc*, 2016.
[55] L. Yang, et al., "Crowd-empowered privacy-preserving data aggregation for mobile crowdsensing," in *Proceedings of ACM MobiHoc*, 2018.
[56] Z. Zhang, et al., "REAP: An Efficient Incentive Mechanism for Reconciling Aggregation Accuracy and Individual Privacy in Crowdsensing," *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp.2995–3007, May 2018.
[57] Z. Huang and S. Kannan, "The exponential mechanism for social welfare: Private, truthful, and nearly optimal," in *Proceedings of the IEEE FOCS*, 2012.
[58] L. Blumrosen and N. Nisan, *Combinatorial auctions*. New York, USA: Cambridge University Press, 2007.
[59] "Nyc311 opendata." [Online]. Available: https://data.cityofnewyork.us/dataset/311-Service-Requests-From-2011/fpz8-jqf4
[60] Nyc311 mobile app. [Online]. Available: http://www1.nyc.gov/nyc-resources/service/5460/nyc311-mobile-app
[61] Cplex optimizer. [Online]. Available: https://www.ibm.com/analytics/data-science/prescriptive-analytics/cplex-optimizer
[62] C. Lund and M. Yannakakis, "On the hardness of approximating minimization problems," *Journal of the ACM*, vol. 41, no. 5, pp. 960–981, 1994.
[63] Myerson and B. Roger , "Optimal auction design," *Mathematics of operations research* , vol. 6, pp. 58–73, May 1981.

Wenqiang Jin received he B.E. and M.E. in Electronic Engineering from Chongqing University of Posts and Telecommunications, China, in 2011 and 2017, respectively. He is currently an Ph.D. student in the Department of Computer Science and Engineering, The University of Texas at Arlington. His research interests include mobile crowd sensing and IoT security.

Mingyan Xiao received her M.S. degree from National University of Defense Technology in 2017, and B.E. from Nanjing University of Aeronautics and Astronautics in 2014, respectively. She is currently a Ph.D. student in the department of Computer Science at The University of Texas at Arlington. Her recent research interests are in the area of resource allocation and management in wireless networks, data-driven security and privacy.



Yang Lei received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 2005 and 2008, respectively, and the Ph.D. degree from the School of Electrical Computer and Energy Engineering, Arizona State University, Tempe, AZ, USA, in 2012. He was a Postdoctoral Scholar with Princeton University, Princeton, NJ, USA, and an Assistant Research Professor with the School of Electrical Computer and Energy Engineering, Arizona State University. He is currently an Assistant Professor with the Department of Computer Science and Engineering, University of Nevada, Reno, NV, USA. His research interests include big data analytics, edge computing and its applications in IoT and 5G, stochastic optimization and modeling in smart cities and cyber-physical systems, data privacy and security in crowdsensing, and optimization and control in mobile social networks. He was a recipient of the Best Paper Award Runner-up at the IEEE INFOCOM 2014. He is currently associate editor for IEEE Access.

Linke Guo received the BE degree in electronic information science and technology from the Beijing University of Posts and Telecommunications in 2008. He received the MS and PhD degrees in electrical and computer engineering from the University of Florida in 2011 and 2014, respectively. From August 2014 to August 2019, he was an assistant professor at the Department of Electrical and Computer Engineering, Binghamton University, State University of New York. Starting from August 2019, he has been an assistant professor with Department of Electrical and Computer Engineering, Clemson University.

Ming Li received the B.E. degree in Electrical Engineering from Sun Yat-sen University, China, in 2007, the M.E. degree in Electrical Engineering from Beijing University of Posts and Communications, China, in 2010, and the Ph.D. degree in Electrical and Computer Engineering from Mississippi State University, Starkville, in 2014, respectively. She is currently an assistant professor in the Department of Computer Science and Engineering, The University of Texas at Arlington. Her research interests include mobile computing, internet of things, security, and privacy-preserving computing. Her work won Best Paper Awards in Globecom 2015 and DASC 2017, respectively. She received the NSF CAREER Award in 2020 and is a member of the IEEE and the ACM.