Privacy Considerations in Participatory Data Collection via Spatial Stackelberg Incentive Mechanisms

Jing Yang Koh¹ · Gareth W. Peters² · Ido Nevat³ · Derek Leong⁴

Received: 20 October 2019 / Revised: 20 October 2019 / Accepted: 25 May 2020 /

Published online: 9 July 2020

© The Author(s) 2020

Abstract

Mobile crowd sensing is a widely used sensing paradigm allowing applications on mobile smart devices to routinely obtain spatially distributed data on a range of user attributes: location, temperature, video and audio. Such data then typically forms the input to application specific machine learning tasks to achieve objectives such as improving user experience, targeting geo-localised query based searches to user interests and commercial aspects of targeted geo-localised advertising. We consider a scenario in which the sensing application purchases data from spatially distributed smartphone users. In many spatial monitoring applications, the crowdsourcer needs to incentivize users to contribute sensing data. This may help ensure collected data has good spatial coverage, which will enhance quality of service provided to the application user when used in machine learning tasks such as spatial regression. Privacy considerations should be addressed in such crowd sensing applications, and an incentive offered to “privacy-concerned” users to contribute data. A novel Stackelberg incentive mechanism is developed that allows workers to specify their location whilst satisfying their location privacy requirements. The Stackelberg and Nash equilibria are explored and an algorithm to demonstrate the approach is developed for a real data application.

Keywords Incentive mechanism design · Stackelberg game · Location privacy · Mobile crowd sensing · Privacy

Mathematics Subject Classification (2010) 91 Game theory, economics, finance and ...
91Bxx · 91Axx · 62 Statistics · 62Cxx · 68 Computer Science · 68Wxx

✉ Gareth W. Peters
garethpeters78@gmail.com

¹ Department of Electrical, Computer Engineering, National University of Singapore, Singapore 117576, Singapore

² Department of Actuarial Mathematics and Statistics, Heriot-Watt, Edinburgh, EH14 4AS UK

³ TUMCREATE Ltd, 1 CREATE Way, CREATE Tower, #10-02 Singapore 138602, Singapore

⁴ Institute for Infocomm Research (I2R), Agency for Science, Technology Research (A*STAR), Singapore 138632 Singapore

1 Introduction

Mobile crowd sensing is becoming a widely used sensing paradigm as it allows applications on smart devices to routinely and efficiently obtain spatially distributed data on a diverse range of user attributes. Such data then typically forms the basis of input data in a variety of machine learning tasks that could have objectives such as improving user experience of applications they run on their smart phone or mobile device, targeting query based searches to their particular interests and geographic location as well as the commercial aspect of targeted advertising.

Previous studies on crowd sourcing data in machine learning contexts, such as Shah et al. (2015) and Shah and Zhou (2016), focused on aspects of provision of labeled training data that is facilitated by crowdsourcing frameworks. In such works, the authors focused on interesting questions pertaining to incentive mechanisms and developed voting schemes to improve aspects of the data collection in such frameworks to address practical concerns such the quality of the crowd sourced labels. They argued that such settings were of relevance in two contexts. Settings in which the application users (typically termed workers) were performing crowdsourced data labeling when in fact they were not experts on the labeling task or alternatively when the application interface is not adequately designed to ensure that workers were able to convey their knowledge accurately on the labeling task.

In this work we consider different aspects of the crowdsourcing data provision framework that were not previously studied. However, we believe they are of direct practical relevance to numerous applications in this emerging area of crowdsourced data provision for machine learning algorithms. We consider the situation in which the sensing application may incentivise or “buy” sensor data from spatially distributed mobile smartphone users (workers) instead of deploying their own sensor networks. In such contexts, the geo-location of the sourced data is directly relevant to the machine learning task that the application will consider when providing the content to the users of the application. Furthermore, we consider an important aspect not previously addressed related to the ability to incentivise crowd sourced data in the presence of workers who are privacy aware and therefore require greater monetary incentives to forgo or relax their privacy concerns when supplying data to the crowd sourcing application.

The spatial aspect of such problems, where the geo-location of sourced data is of direct relevance, is previously under explored but is relevant to many unsupervised or semi-supervised machine learning application contexts. For instance, we explore a setting in which one may wish to perform a statistical task of say Gaussian process spatial field regression or estimation using crowdsourced data. Such tasks can then provide estimates of application specific information to a user at the location of the smartphone, based on a statistical models calibrated to previously sourced spatially distributed data of relevance. Numerous examples come to mind for such applications including localized climate information, traffic density information, local pollution levels etc.

Such mobile crowd sensing platforms Guo et al. (2015) have also become an emerging sensing paradigm in the age of the Internet of Things (IoT) that replaces fixed sensing infrastructure and removes its deployment and maintenance costs. The interface of such IoT architectures and statistical machine learning tasks is an exciting new area beginning to emerge, which will require ideas such as those addressed in this manuscript to facilitate data sourcing.

As noticed in Murakami et al. (2016), such sensing platforms can also be practically useful to exploit the growth of mobile smartphone ownership and tap into smartphone users to contribute sensing data that are easily obtained via the available sensing capabilities of their

smartphones. This allows the sensing platform to then utilise this spatially distributed data to perform important practical tasks such as to estimate some statistics of a spatial event or to conduct spatial regression. Other examples of mobile crowd sensing applications for spatial monitoring include environmental temperature monitoring Overeem et al. (2013), traffic monitoring Thiagarajan et al. (2009) or earthquake detection in early warning systems Minson et al. (2015). All of these aforementioned application domains have benefited from the fact that mobile crowdsourced data has improved the *spatial coverage* of the collected dataset and therefore improved the resolution of spatial regression model estimations. In our framework, we consider the aspect of spatial coverage as one of the main objectives of an incentive mechanism used by spatial monitoring applications.

Additionally, current privacy-preserving works such as Nissim et al. (2012), Yang et al. (2013), and Singla and Krause (2013) have attempted to address the user *location privacy* problem in the crowd sensing domain. This is because the privacy issues can easily deter potential users from participating, which in turn reduces both the amount of available data as well as spatial coverage of the data used by the crowdsourcer.

1.1 Location-Privacy Preservation and Incentivised Crowdsourced Data

Existing privacy-preserving works that offer location privacy via location or data perturbation are not directly applicable to crowd sensing applications that require specific and true locations. For example, it would be unacceptable for a traffic monitoring application if there was a traffic congestion in road X, but due to location or data perturbation, another road Y or a non-congested status was reported respectively. Thus, it is vital for incentive models to address the spatial coverage and location privacy issues concurrently.

The most common incentive mechanism is the auction game, see examples in Yang et al. (2012), Nissim et al. (2012), and Singla and Krause (2013) and (Restuccia et al. 2016). In such settings the smartphone workers submit bids (e.g., prices or efforts in Luo et al. (2016b)) for their sensing data while the crowdsourcer selects the set of workers with the lowest bids and pays them accordingly.

An alternative and fundamentally different game-theoretic approach to model the incentive problem is the Stackelberg leader-follower game where the crowdsourcer (leader) first decides on the total reward to pay workers (i.e., the leader's budget constraint) while the workers (followers) then individually decide on the amount of data to contribute, see Yang et al. (2012) and Luo et al. (2016a).

However, different from the system models used in Yang et al. (2012) and Luo et al. (2016a) where the crowdsourcer first informs the workers of the total reward offered (Fig. 1). Then subsequently, the workers compute the optimal data to contribute (which may not be intuitive or practically feasible for the smartphone users in practice), we allow the

Fig. 1 Privacy model of users: privacy-concerned users can declare a cloaking region that contains their true location instead of providing fine-grained location information to the crowdsourcer

crowdsourcer to specify the exact reward offered to each individual worker (see stage 2 in Fig. 2). In this way, the workers can be assured that the offered rewards are always their optimal (Nash equilibrium) solution and there is no need to share all the information on the workers' costs and privacy preferences between other competing workers.

Privacy-aware incentive mechanisms were considered in Nissim et al. (2012), Yang et al. (2013), and Singla and Krause (2013) which used data perturbation or dummy locations Kido et al. (2005) to protect location privacy. However, the methods may not be applicable to many spatial monitoring applications where a dataset with perturbed data or dummy location may trigger a false alarm and make the application unreliable. In this paper, we allow workers to obfuscate their precise location information by declaring a coarse-grained region (see Fig. 1) that encompasses their true location (similar to the location cloaking principle in Cheng et al. (2006) and Andrés et al. (2013)). This step is important as privacy concerns can deter potential workers from participating in the crowd sensing activity. In addition, the work in Lin et al. (2014) found that smartphone users were more willing to provide coarse-grained location information than fine-grained information. Hence, our Stackelberg model may still buy data (albeit with lower quality location information) from privacy-sensitive workers to improve the spatial coverage of the collected dataset. Furthermore, the cloaking region technique is more practical for real-world applications that require reliable information. Although the location information of the workers may be *imprecise* due to the location cloaking, but it is still *accurate* as there is no data perturbation or dummy locations involved.

Additionally, existing Stackelberg models assume that workers sell undifferentiated goods (data) and do not consider the quality (e.g., granularity) of each worker's data and the workers' contribution to the spatial coverage of the collected dataset. However, adding location information increases the computational complexity of the problem and thus, Feng et al. (2014) proposed an auction-based approximation algorithm to assign the workers' sensing tasks. The authors assumed that the sensing platform periodically publishes sensing tasks for specific locations of interest and did not explicitly address the issue of improving the spatial coverage of the collected dataset in general.

1.2 Our Contributions

We demonstrate that an appropriate incentive mechanism to model the hierarchical relationship between the crowdsourcer and the smartphone users is the *Stackelberg* (*leader-*

Fig. 2 Interaction model between the crowdsourcer and smartphone users (workers). The user profile consists of the user's sensing cost incurred per unit of data c_i , location granularity ρ_i , and partitioned region l_i

follower) game model used in Yang et al. (2012) and Luo et al. (2016a). In the Stackelberg model, the crowdsourcer (leader) commits a reward strategy that is observed by the smartphone users (followers) who then strategize the amount of data to sell. However, existing Stackelberg incentive models simply select user data independently of their physical location Feng et al. (2014) and do not attempt to improve the spatial coverage of the dataset.

Therefore, we propose extending the existing Stackelberg incentive models to include the *privacy-awareness* property as well as to improve the *spatial coverage* of the collected dataset. Our model allows privacy-sensitive users to submit *coarse-grained* (or *quantized*) *location* information which could still be useful to the crowdsourcer. We then study the properties of the proposed Stackelberg incentive model analytically and present efficient algorithmic solutions. Our proposed Stackelberg incentive model does not require a trusted third party for privacy and can protect users against a crowdsourcer who cannot be trusted to anonymize the smartphone users' location information. The current work extends our preliminary work in Koh et al. (2017), by accommodating bounds in the workers' sensing data, analyzing the existence, uniqueness and Pareto efficiency of the Stackelberg equilibrium, and conducting a real-world sensing case study to demonstrate the practicality of the proposed solution.

The main contributions of the paper are as follows:

- We propose a novel spatial privacy-aware Stackelberg incentive scheme that allows privacy-sensitive mobile smartphone users to quantize their location information using cloaking regions. Our proposed solution also seeks to improve the spatial coverage of the collected dataset.
- We prove that under our Stackelberg incentive mechanism, we are able to develop a unique Nash equilibrium for the Followers game, and obtain a closed-form expression for each mobile smartphone user's optimal data contribution.
- We also prove the existence of a Stackelberg equilibrium when the crowdsourcer (leader) imposes constraints on the minimum and maximum amount of data contribution from each user and we further derive sufficient conditions for the Stackelberg equilibrium to be Pareto efficient.
- We demonstrate via simulations using a real-world sensing dataset that the proposed spatial privacy-preserving Stackelberg game produces a mechanism that for a common cost, will produce greater spatial diversity in crowdsourced data, leading to better model predictive performance in applications such as spatial field reconstruction compared to two other coverage-maximizing schemes.

2 Problem Formulation and Analysis

In this paper, we address the following problem statement: suppose there is a crowdsourcer (who is the data buyer and leader on the demand side) who aims to buy sensing data from mobile smartphone users (who are the workers and followers on the supply side) for applications such as spatial monitoring; we aim to design an incentive mechanism such that the collected dataset: (i) has good *spatial coverage*, and (ii) is *location privacy-preserving* for the workers. We first present our system model and the proposed Stackelberg incentive model before studying the proposed game analytically. We extend the results in (Yang et al. 2012, Section 3) to take into account the workers' location granularities and regions. The table of notation is given in Table 1.

Table 1 Notation

c_i	Sensing cost incurred per unit of data by worker i where $c_i \in (0, \bar{c}]$.
$\mathbf{c}_{-i}$	Sensing cost incurred per unit of data by all workers other than worker i .
ρ_i	Location granularity of worker i where $\rho_i \in [\underline{\rho}, \bar{\rho}]$.
ρ_{-i}	Location granularities of all workers other than worker i .
l_i, l'_i	Partitioned region and cloaking region of worker i respectively where $l'_i \in l_i$.
t_i	Amount of data contributed by worker i .
$\mathbf{t}_{-i}$	Amount of data contributed by all workers other than worker i .
R_l	Sum of rewards allocated to workers in region l (similarly, R_{l_i} refers to the sum of rewards allocated to worker i 's region l_i).
α_i	System parameter for crowdsourcer.
β	Diminishing returns parameter for crowdsourcer.
U_{CS}, u_i	Utility function of crowdsourcer and worker i respectively.
$\mathcal{I}$	Set of all workers.
N_l	Number of workers in region l .
$\mathcal{Q}_l$	Set of participating workers in region l (similarly, $\mathcal{Q}_{l_i}$ refers to the set of participating workers in the region l_i where worker i is located).
$\mathcal{J}_l$	Set of participating workers i in region l with $t_i = \bar{t}$.
$\mathcal{K}_l$	Set of participating workers i in region l with $t_i < \bar{t}$.

2.1 Model and Assumptions

We consider a crowd sensing system that consists of a set of $\mathcal{I} = \{1, \dots, N\}$ workers and a single crowdsourcer who partitions the entire spatial area of interest into a set of L regions denoted by $\mathcal{L}$. We assume that the workers are rational and non-cooperative, i.e., each worker maximizes its own utility. Each worker $i \in \mathcal{I}$ has its own sensing cost per unit of data $c_i \in (0, \bar{c}]$, location granularity $\rho_i \in [\underline{\rho}, \bar{\rho}]$, cloaking region l'_i (which may be of different granularity for each worker) and its corresponding region l_i (defined by the crowdsourcer) where $l'_i \subseteq l_i \subseteq \mathcal{L}$. The interaction model between the crowdsourcer and the mobile smartphone users is illustrated in Fig. 2. Essentially, the crowdsourcer collects the worker profiles (c_i, ρ_i, l_i) and selects the optimal set of workers that maximizes its utility. The crowdsourcer then offers the selected workers a reward in exchange for some specified amount of sensing data. Next, the selected workers (which we refer to as participating workers) will proceed to collect their sensing data and transmit it along with their cloaking region l'_i to the crowdsourcer and receive their rewards in return. Note that there are two types of regions l_i and l'_i in our model: l_i is the initial coarse-grained partitioned region defined by the crowdsourcer and l'_i is worker i 's cloaking region, which is submitted only when the worker is selected.

We assume that the granularity of the worker i 's submitted cloaking region l'_i is proportional to its location granularity parameter ρ_i , since a privacy-sensitive worker (with low ρ_i) is likely to provide only coarse-grained location information, while a privacy-insensitive worker (with high ρ_i) is more likely to provide finer-grained location information. Hence, the parameter ρ_i allows the crowdsourcer to differentiate between workers with different levels of sensitivity to privacy and preserve the privacy of unselected workers since they only reveal more information on their locations when they are selected. Note that the workers can anonymize their precise location information via cloaking regions with area inversely

proportional to ρ_i and do not rely on the crowdsourcer for anonymization. In practice, this can be implemented in the crowdsourcing software on the workers' smartphone devices to allow each worker to select cloaking regions with different (possibly discrete) location granularities.

We model the incentive mechanism as a *Stackelberg game*, which consists of the crowdsourcer (data buyer) as the *leader* and the N smartphone users (workers and data contributors) as the *followers*. The crowdsourcer acts first and commits a reward strategy while the workers subsequently choose their best responses after observing the crowdsourcer's strategy. The strategy of the crowdsourcer is the reward for each partitioned region $\mathbf{R} = (R_1, \dots, R_L)$ and the strategy of worker i is the amount of data contributed $t_i \geq 0$. The crowdsourcer only optimizes the reward R_l allocated to each partitioned region $l \in \mathcal{L}$ and subsequently offers each participating worker i in the partitioned region l a fraction of R_l depending on the proportion of data contribution, weighted by the location granularity:

$$\text{Worker } i\text{'s offered reward} = \frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{l_i}, \quad (1)$$

where $\mathcal{Q}_{l_i}$ is the set of participating workers i in partitioned region $l_i \in \mathcal{L}$ with $t_i > 0$ and we assume $|\mathcal{Q}_{l_i}| > 1$. A similar reward function has also been used in (Yang et al. 2012).

2.2 Utility Functions for Crowdsourcer and Workers

In this section we introduce the formulations for the utility functions for each party in the crowdsourcing data mechanism.

Crowdsourcer: We define the utility function of the crowdsourcer to be $U_{CS}(\mathbf{R}; \mathbf{t}) = \sum_{i \in \mathcal{I}} f_d(t_i, l_i, \rho_i)$, where $f_d(t_i, l_i, \rho_i)$ is a function of worker i 's amount of data contributed t_i , partitioned region l_i , and its location granularity ρ_i . For simplicity, we let $f_d(t_i, l_i, \rho_i) = \alpha_i t_i^\beta$ where $\alpha_i > 0$ is a function of (l_i, ρ_i) , and the parameter $0 < \beta < 1$ is used to control the rate of diminishing returns on each worker's data. The power function was also used in (Powell and Batt 2008) to model diminishing returns. Thus, the crowdsourcer's utility function is given by:

$$U_{CS}(\mathbf{R}; \mathbf{t}) = \sum_{i \in \mathcal{I}} \alpha_i t_i^\beta. \quad (2)$$

To increase the coverage area of the collected dataset, the crowdsourcer can assign a higher α_i value to workers located at less populated regions l_i . In addition, a higher α_i value can be assigned to workers that provide finer location granularity ρ_i . The α_i parameter allows the crowdsourcer to differentiate between the quality (e.g., spatial coverage area and the granularity of the location information) of each worker's data.

Workers: We define the utility function of worker i to be the amount of reward received from the crowdsourcer as defined in Eq. 1 minus the cost incurred for obtaining the data:

$$u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) = \frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{l_i} - c_i t_i, \quad (3)$$

where $\mathbf{t}_{-i}$ is a vector of the amount of data contributed by all workers except worker i .

2.3 Stackelberg Game Formulation

Given that the crowdsourcer wants to achieve a large coverage area for its dataset while satisfying a budget constraint R^{budget} , a minimum amount of reward $R_l^{\min} > 0$ for each region l (this allows the crowdsourcer to specify more important regions), and a maximum amount of reward allocation $R_l^{\max} > 0$ for each region l (which may be set to ∞), it solves the following optimization problem:

Problem 1 ($\alpha, \beta, \mathbf{R}^{\min}, \mathbf{R}^{\max}, R^{\text{budget}}$):

$$\begin{aligned} & \underset{\mathbf{R}}{\text{maximize}} && \sum_{i \in \mathcal{I}} \alpha_i t_i^\beta \\ & \text{subject to} && R_l^{\min} \leq R_l \leq R_l^{\max}, \quad \forall l \in \mathcal{L}, \\ & && \sum_{l \in \mathcal{L}} R_l \leq R^{\text{budget}}, \end{aligned} \quad (4)$$

where t_i is the optimal solution to Problem 2.

Each worker i solves the following optimization problem:

Problem 2 ($i, \mathbf{t}_{-i}, \rho_i, \rho_{-i}, R_{li}, c_i$):

$$\begin{aligned} & \underset{t_i}{\text{maximize}} && \frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{li} - \\ & && c_i t_i, \\ & \text{subject to} && t_i \geq 0. \end{aligned} \quad (5)$$

Problems 1 and 2 form the Stackelberg game and our goal is to find the Stackelberg equilibrium point(s) where neither the crowdsourcer nor the workers have incentive to deviate. A Stackelberg equilibrium (see Definition 1) is a subgame-perfect Nash equilibrium such that no player can improve its utility by unilaterally deviating from its strategy.

Definition Stackelberg Equilibrium Let $\mathbf{R}^*$ be the optimal solution for the crowdsourcer, obtained by solving Problem 1, and $\mathbf{t}^*$ be the optimal solution for the workers, obtained by solving Problem 2. The strategy profile $(\mathbf{R}^*, \mathbf{t}^*)$ is a Stackelberg equilibrium for the proposed Stackelberg incentive model if the following conditions are satisfied for any $(\mathbf{R}, \mathbf{t})$ where:

$$\begin{aligned} & U_{CS}(\mathbf{R}^*; \mathbf{t}^*) \geq U_{CS}(\mathbf{R}; \mathbf{t}^*), \\ & u_i(t_i^*; \mathbf{t}_{-i}^*, \mathbf{R}^*) \geq u_i(t_i; \mathbf{t}_{-i}^*, \mathbf{R}^*), \quad \forall i \in \mathcal{I}. \end{aligned}$$

We apply the backward induction method to analyze the proposed Stackelberg incentive model. First, we start with the Followers game (a non-cooperative game played by all the workers) and study the predicted best response t_i^* (solution of Problem 2) for each worker i as a function of the reward R_{li} offered by the crowdsourcer and the strategies of the other workers $\mathbf{t}_{-i}$. Subsequently, we analyze the best response of the crowdsourcer in Problem 1.

2.4 Nash Equilibrium of Followers game

We first consider the Followers game given by the triplet $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$ where $\mathcal{I}$ is the player set of N workers and u_i is the utility function of worker i . By Lemma 2, there

exists a unique Nash equilibrium point in the Followers game. Next, we derive the unique Nash equilibrium of the Followers game in Theorem 3.

Lemma A unique Nash equilibrium exists in the Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$.

Proof: See Appendix 1.

Theorem The Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$ has a unique Nash equilibrium given by the following closed-form expression:

$$t_i^* = \begin{cases} \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_{l_i}| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i} & \text{if } i \in Q_{l_i}, \\ 0 & \text{otherwise,} \end{cases} \quad (6)$$

where Q_{l_i} is the set of participating workers in region l_i .

Proof: See Appendix 2.

We say that worker i is a participating worker if $t_i^* > 0$. Hence, from Eq. 6, all participating workers $i \in Q_{l_i}$ should satisfy the constraint:

$$\begin{aligned} & \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_{l_i}| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i} > 0 \\ & \iff \left(1 - \frac{(|Q_{l_i}| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right) > 0 \\ & \iff \frac{c_i}{\rho_i} < \frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1}. \end{aligned} \quad (7)$$

Subsequently, we will make use of this constraint in Algorithm 1 (modified from (Yang et al. 2012, Algorithm 1)), which computes the Nash equilibrium solution for the Followers game.

The optimal t_i^* for the workers is given by the Nash equilibrium solution Eq. 6 of the Followers game. However, the expression in Eq. 6 requires knowledge of the set of participating workers Q_{l_i} , the sensing costs c_i , and the location granularity ρ_i of all workers $i \in \mathcal{I}$. Hence, we propose Algorithm 1 which makes use of Eq. 7 to greedily compute Q_{l_i} and solve for the optimal t_i^* of the workers (in terms of R_{l_i}). The values of t_i^* can then be substituted into Problem 1 to solve for $\mathbf{R}^*$.

Lemma Assume that there are at least two workers in each region l . Algorithm 1 selects the set of participating workers Q_l that achieves the unique Nash equilibrium solution of the Followers game.

Proof: See Appendix 3.

Theorem 5 proves the correctness of Algorithm 1 while Lemma 4 simply states that the Q_l set computed in Algorithm 1 is correct.

Theorem Assuming that there are at least two workers in each region l , and the set of $\mathcal{Q}_l$ from Lemma 4, then Algorithm 1 outputs the unique Nash equilibrium solution of the Followers game.

Proof: See Appendix 4.

2.5 Stackelberg Equilibrium

Using the analytical result Eq. 6 for the Followers game, the crowdsourcer can optimize its reward strategy $\mathbf{R}$ efficiently by substituting the analytical result into its utility function in Eq. 2 to obtain:

$$U_{CS}(\mathbf{R}; \mathbf{t}) = \sum_{i \in \mathcal{I}} \alpha_i (\tau_i R_{l_i})^\beta, \quad (8)$$

$$\text{where } \tau_i = \frac{(|\mathcal{Q}_{l_i}| - 1)}{\rho_i \sum_{j \in \mathcal{Q}_{l_i}} \frac{c_j}{\rho_j}} \left(1 - \frac{(|\mathcal{Q}_{l_i}| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c_j}{\rho_j}} \right).$$

Although it is not trivial to obtain a closed-form expression for $\mathbf{R}$ that maximizes Eq. 8 while satisfying the constraints in Eq. 4, we show in Theorem 6 that there exists a unique Stackelberg equilibrium which results in a stable equilibrium strategy profile. This allows the crowdsourcer to uniquely predict the behaviors of the workers and efficiently compute its optimal $\mathbf{R}^*$. Both Theorems 3 and 6 extend (Yang et al. 2012, Theorem 2) to the case where the workers' location granularities ρ and regions l are also considered.

Theorem The proposed Stackelberg incentive model has a unique Stackelberg equilibrium.

Proof Recall from Theorem 3 that the Followers game has a unique Nash equilibrium point. It can be easily shown that the best strategy set of the crowdsourcer is convex and compact since the domain of $\mathbf{R}$ is a Cartesian product of closed intervals, and U_{CS} is continuous in $\mathbf{R}$. Hence, we need to show the strict concavity of U_{CS} to conclude that there exists a unique maximum point. The partial derivatives of U_{CS} with respect to R_l are as follows:

$$\begin{aligned} \frac{\partial U_{CS}}{\partial R_l} &= \sum_{i: l_i=l} \frac{\alpha_i \beta (\tau_i R_l)^\beta}{R_l}, \\ \frac{\partial^2 U_{CS}}{\partial R_l^2} &= \sum_{i: l_i=l} \frac{\alpha_i (\beta - 1) \beta (\tau_i R_l)^\beta}{R_l^2}, \quad \frac{\partial^2 U_{CS}}{\partial R_l \partial R_k} = 0, \end{aligned} \quad (9)$$

where τ_i is given in Eq. 8. Note that $\frac{\partial^2 U_{CS}}{\partial R_l^2} < 0$ as we assume $\alpha_i > 0$ for all $i \in \mathcal{I}$ and $0 < \beta < 1$. Since the Hessian matrix of U_{CS} is a diagonal matrix, its eigenvalues are given by $\frac{\partial^2 U_{CS}}{\partial R_l^2}$, which are all strictly negative. This implies that the Hessian matrix is negative definite and thus, U_{CS} is strictly concave in $\mathbf{R}$. $\square$

Using the results from Theorem 5 (which gives the closed-form solution to Problem 2), the Stackelberg equilibrium solution can be computed by solving Problem 1, which is a

convex optimization problem. By Theorem 6, there exists a unique reward strategy $\mathbf{R}^*$ that maximizes the crowdsourcer's utility in Problem 1. This means that $\mathbf{R}^*$ can be efficiently computed using well-known interior point methods.

3 Analysis of Optimal Declared Sensing Costs

In this section, we study the strategies of the workers in selecting their optimal sensing costs to declare to the crowdsourcer. This is essential as each worker may declare a false sensing cost to maximize its utility. For this section, we let the declared sensing cost for each worker i be denoted by c'_i and its true sensing cost be denoted by c_i . Note that the true sensing cost c_i is unknown to the crowdsourcer and other workers $j \in \mathcal{Q}_{l_i}$ where $j \neq i$.

Suppose that each worker i may lie about its sensing cost to maximize its utility u_i , in such a dishonest reporting scenario, the optimization problem previously specified in Problem 2 is replaced with that considered in Problem 3 below. The interaction model previously shown in Fig. 2 needs to be updated as each worker i requires additional information on $\mathcal{Q}_{l_i}$ and $\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}$ to solve Problem 3. One way for the workers to obtain the required information would be for the crowdsourcer to provide a platform for the workers to view the required information for a fixed time period. If the workers can declare or modify their c'_i values during the fixed time period, then they can achieve the Nash equilibrium solution given in Theorem 9.

Problem 3 ($i, c_i, \mathbf{c}_{-i}, \rho_i, \rho_{-i}, R_{l_i}$):

$$\begin{aligned} & \underset{c'_i}{\text{maximize}} \quad \frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{l_i} - c_i t_i, \\ & \text{subject to} \quad 0 \leq c'_i \leq \bar{c}, \end{aligned} \quad (10)$$

where $t_i = \frac{|\mathcal{Q}_{l_i}|-1}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}} \left(1 - \frac{(|\mathcal{Q}_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i}$ if $i \in \mathcal{Q}_{l_i}$, and 0 otherwise, according to Eq. 6.

We first derive the expression for the worker i 's utility in terms of its declared sensing cost and the declared sensing costs of the other participating workers j in its region

l_i . Using Algorithm 1, the crowdsourcer offers to buy $t_i = \frac{|\mathcal{Q}_{l_i}|-1}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}} \left(1 - \frac{(|\mathcal{Q}_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i}$

amount of data from each worker i according to Eq. 6. From Eq. 26, we have the expression: $\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j = \frac{|\mathcal{Q}_{l_i}|-1}{\sum_{j \in \mathcal{Q}_{l_i}} \frac{c'_j}{\rho_j}} R_{l_i}$. We substitute the expressions for t_i and $\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j$ into the

utility of worker $i \in \mathcal{Q}_{l_i}$ in Eq. 3 to obtain:

$$\begin{aligned}
u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) &= \frac{\frac{|Q_{l_i}|-1}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \left(1 - \frac{(|Q_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i}}{\frac{|Q_{l_i}|-1}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} R_{l_i}} R_{l_i} - c_i \left(\frac{|Q_{l_i}|-1}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) \left(1 - \frac{(|Q_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i} \\
&= \left(1 - \frac{(|Q_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) R_{l_i} - \left(\frac{c_i}{\rho_i} \right) \left(\frac{|Q_{l_i}|-1}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) \left(1 - \frac{(|Q_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) R_{l_i} \\
&= \left(1 - \frac{(|Q_{l_i}|-1) \frac{c'_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) \left(1 - \frac{(|Q_{l_i}|-1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c'_j}{\rho_j}} \right) R_{l_i}.
\end{aligned} \tag{11}$$

Next, we have the following Theorem 9 on the Nash equilibrium solution of the Followers game when the workers can optimize their declared sensing costs.

Lemma Suppose that worker $i \in Q_l$ declares a sensing cost $c'_i < c_i$, then it always incurs a negative utility $u_i < 0$ if $\frac{c_i}{\rho_i} > \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$ and gets a positive utility $u_i > 0$ if $\frac{c_i}{\rho_i} < \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$.

Proof Suppose worker i is not in the set of participating workers Q_l if it does not lie about its sensing cost, i.e., $\frac{c_i}{\rho_i} \geq \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$. To be in the set of participating workers Q_l , the worker i can declare a sensing cost $c'_i < c_i$ such that $\frac{c'_i}{\rho_i} < \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$. In order for its utility to be negative, i.e., $u_i < 0$, we have the following necessary condition from Eq. 11: $\frac{c_i}{\rho_i} > \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$.

Similarly, for $u_i > 0$, we have the following necessary condition from Eq. 11: $\frac{c_i}{\rho_i} < \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$.

Suppose worker i is in the set of participating workers Q_l if it does not lie about its sensing cost, i.e., $\frac{c_i}{\rho_i} < \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$. In order for its utility to be negative, i.e., $u_i < 0$, we have the

following necessary condition from Eq. 11: $\frac{c_i}{\rho_i} > \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$. Similarly, for $u_i > 0$, we have the following necessary condition from Eq. 11: $\frac{c_i}{\rho_i} < \frac{\sum_{j \in Q_l} \frac{c'_j}{\rho_j}}{|Q_l|-1}$. Hence, the proof is complete. $\square$

Lemma Given the set of participating workers Q_l , the worker $i \in Q_l$ can maximize its

utility u_i if it declares a sensing cost of $c'_i = \frac{|\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} - \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2}{(|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}} \rho_i$.

Proof: See Appendix 5.

Theorem Assuming that each worker $i \in \mathcal{Q}_l$ optimizes its sensing cost c'_i using Lemma 8, then the Nash equilibrium solution of the Followers game occurs when

$$\frac{c'_i}{\rho_i} = \frac{|\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} - \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2}{(|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}}, \forall i \in \mathcal{Q}_l. \quad (12)$$

Proof By Lemma 7, we have the constraint $\frac{c_i}{\rho_i} < \frac{\sum_{j \in \mathcal{Q}_l} \frac{c'_j}{\rho_j}}{|\mathcal{Q}_l| - 1}$ for all workers $i \in \mathcal{Q}_l$ to ensure that all participating workers $i \in \mathcal{Q}_l$ have a positive utility, i.e., $u_i > 0$. By Lemma 8, the worker $i \in \mathcal{Q}_l$ can maximize its utility u_i if it declares a sensing cost of

$$c'_i = \frac{|\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} - \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2}{(|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}} \rho_i.$$

□

Remarks: If we assume that the workers $i \in \mathcal{Q}_l$ have a fixed time period to declare or modify their c'_i values and that the crowdsourcer provides a platform for each individual worker to observe the total declared sensing cost c'_j and location granularities ρ_j of all other workers $j \in \mathcal{Q}_l$ in the same region, then the equilibrium point in Eq. 12 can be achieved.

4 Adding Bounds on the Amount of Workers' Data and Achieving Pareto Efficiency

In this section, we study how the crowdsourcer is able to introduce bounds on the workers' data contributions and how it can achieve Pareto efficiency for the Stackelberg equilibrium.

4.1 Bounds on the Amount of Contributed Data t_i

In a practical real-world crowdsourcing application, the crowdsourcer may impose (lower and upper) bounds on the amount of data contribution from each worker i due to various reasons. For example, it may not be useful to the crowdsourcer if each worker only contributes a small amount of data, hence we consider the imposition of a lower bound $\underline{t}_\rho$ on

the workers' data contribution. In the other extreme, to maximize the spatial coverage of the obtained data, it may not be useful to the crowdsourcer if there are a small number of workers who are monopolies. Hence, we also consider the imposition of an upper bound $\overline{t\rho}$ on the workers' data contribution. For simplicity, we assume that the same bounds $\underline{t\rho}$ and $\overline{t\rho}$ apply to all participating workers.

Next, we have the following Theorem 10 to introduce the lower and upper bounds on the worker i 's $t_i \rho_i$ values and Theorem 14, which states the sufficient condition for the bounds. Finally, we present Algorithm 2, which computes the bounded Nash equilibrium solution for the Followers game.

Theorem Assume that $\frac{c_i}{\rho_i} \leq 1$ for all $i \in \mathcal{I}$. The crowdsourcer is able to introduce both the lower bound $\underline{t\rho}$ and upper bound $\overline{t\rho}$ in the Stackelberg equilibrium of all participating workers i in region l if the following constraints are satisfied:

$$\frac{c_i}{\rho_i} \leq \frac{\sum_{j \in Q_l} \frac{c_j}{\rho_j}}{|Q_l| - 1} \left(1 - \frac{\underline{t\rho} \sum_{j \in Q_l} \frac{c_j}{\rho_j}}{(|Q_l| - 1)R_l} \right), \quad \forall i \in Q_l,$$

$$R_l^{\max} = \overline{t\rho} \left[\frac{|Q_l| - 1}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_l| - 1) \frac{c_l^M}{\rho_l^M}}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \right) \right]^{-1}, \quad (13)$$

where we let the sensing cost and location granularity of the worker i with the least $\frac{c_i}{\rho_i}$ value in region l be denoted by c_l^M and ρ_l^M respectively.

Proof We first show the following two Lemmas, which are used to prove the theorem:

Lemma Suppose that $\frac{c_i}{\rho_i} \leq 1$ for all $i \in \mathcal{I}$ and Eq. 14 is true. Then the crowdsourcer is able to introduce a lower bound $\underline{t\rho}$ in the Stackelberg equilibrium of all participating workers i in region l .

$$\frac{c_i}{\rho_i} \leq \frac{\sum_{j \in Q_l} \frac{c_j}{\rho_j}}{|Q_l| - 1} \left(1 - \frac{\underline{t\rho} \sum_{j \in Q_l} \frac{c_j}{\rho_j}}{(|Q_l| - 1)R_l} \right), \quad \forall i \in Q_l. \quad (14)$$

We first prove Lemma 11. To show that the Stackelberg solution exists in the constrained region, one must show that $t_i^* \rho_i \geq \underline{t\rho}$ for all $i \in Q_l$. To prove that $t_i^* \rho_i \geq \underline{t\rho}$ for all $i \in Q_l$, we first let $\frac{c_i}{\rho_i} = \frac{\sum_{j \in Q_l} \frac{c_j}{\rho_j}}{|Q_l| - 1} \left(1 - \frac{\underline{t\rho} \sum_{j \in Q_l} \frac{c_j}{\rho_j}}{(|Q_l| - 1)R_l} \right) - \delta_i$ where $\delta_i \geq 0$. We take the solution in Eq. 6 and substitute the $\frac{c_i}{\rho_i}$ expression from above and multiply both sides by ρ_i to obtain:

$$\begin{aligned}
t_i^* \rho_i &= \frac{|\mathcal{Q}_l| - 1}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}} \left(1 - \frac{(|\mathcal{Q}_l| - 1) \left(\frac{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}}{|\mathcal{Q}_l| - 1} \left(1 - \frac{t\rho \sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}}{(|\mathcal{Q}_l| - 1) R_{l_i}} \right) - \delta_i \right)}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}} \right) R_{l_i} \\
&= \frac{|\mathcal{Q}_l| - 1}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}} \left(1 - 1 + \frac{t\rho \sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}}{(|\mathcal{Q}_l| - 1) R_{l_i}} + \frac{(|\mathcal{Q}_l| - 1) \delta_i}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}} \right) R_{l_i} \\
&= t\rho + \left(\frac{|\mathcal{Q}_l| - 1}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j}} \right)^2 R_{l_i} \delta_i.
\end{aligned} \tag{15}$$

Since $\delta_i \geq 0$, we conclude from Eq. 15 that $t_i^* \rho_i \geq t\rho$ for all $i \in \mathcal{Q}_l$.

To verify that the introduction of the lower bound preserves the concavity of the crowdsourcer's Problem 1, we analyze the Hessian matrix of U_{CS} with respect to R_l .

We first apply the chain rule: $\frac{\partial^2 U_{CS}}{\partial R_l^2} = \frac{\partial^2 U_{CS}}{\partial t_i^2} \times \frac{\partial^2 t_i}{\partial R_l^2}$. From Eq. 2, we obtain $\frac{\partial^2 U_{CS}}{\partial t_i^2} = \beta(\beta - 1) \sum_{i \in \mathcal{I}} \alpha_i t_i^{\beta-2}$, where $\beta < 1$. Hence, we have $\frac{\partial^2 U_{CS}}{\partial t_i^2} < 0$. We now attempt to derive

$\frac{\partial^2 t_i}{\partial R_l^2}$. From Eq. 15, we have $t_i \rho_i = t\rho + \left(\frac{|\mathcal{Q}_l| - 1}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j} - \delta_i} \right)^2 R_{l_i}$. If we assume that $\frac{c_i}{\rho_i} \leq 1$

for all $i \in \mathcal{I}$, then the term $\frac{|\mathcal{Q}_l| - 1}{\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j} - \delta_i}$ is inversely proportional to R_{l_i} . This is because

$|\mathcal{Q}_l| \propto R_{l_i}$ according to Eq. 14 and the term $\sum_{j \in \mathcal{Q}_l} \frac{c_j}{\rho_j} - \delta_i$ is proportional to $|\mathcal{Q}_l|$ but increases

at a rate slower than $|\mathcal{Q}_l|$ when $\frac{c_i}{\rho_i} \leq 1$. Therefore, we conclude from Eq. 15, that t_i is a monotonically non-decreasing function of R_l . This implies convexity, i.e., $\frac{\partial^2 t_i}{\partial R_l^2} \geq 0$ given

that t_i is continuous. Since $\frac{\partial^2 U_{CS}}{\partial t_i^2} < 0$ and $\frac{\partial^2 t_i}{\partial R_l^2} \geq 0$, we have $\frac{\partial^2 U_{CS}}{\partial R_l^2} \leq 0$. In Eq. 9, it was shown that $\frac{\partial^2 U_{CS}}{\partial R_l \partial R_k} = 0$. Since $\frac{\partial^2 U_{CS}}{\partial R_l^2} \leq 0$, the Hessian matrix of U_{CS} is negative semidefinite for all $R_l \in \mathbf{R}$. Thus, we conclude that U_{CS} is concave in $\mathbf{R}$.

To introduce the lower bound $t\rho$ constraint in the proposed Stackelberg incentive model, the constraint Eq. 14 should be used instead of constraint Eq. 7 in line 6 of Algorithm 1.

Next, we derive the sufficient conditions for the upper bound $\bar{t}\bar{\rho}$ on the workers' data contributions. For each region l , we let the sensing cost and location granularity of the "cheapest" worker i with the least $\frac{c_i}{\rho_i}$ value be denoted by c_l^M and ρ_l^M respectively. Using the t_i^* expression in Eq. 6, we proceed to derive the maximum $t_i \rho_i$ contributed by the "cheapest" worker in each region l :

$$t_l^M \rho_l^M = \frac{|Q_l| - 1}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_l| - 1) \frac{c_l^M}{\rho_l^M}}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \right) R_l. \quad (16)$$

This leads to the following Lemma 12.

Lemma The crowdsourcer is able to introduce an upper bound $\bar{t}p$ in the Stackelberg equilibrium of all participating workers i in region l if:

$$R_l^{\max} \leq \bar{t}p \left[\frac{|Q_l| - 1}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_l| - 1) \frac{c_l^M}{\rho_l^M}}{\sum_{j \in Q_l} \frac{c_j}{\rho_j}} \right) \right]^{-1}. \quad (17)$$

We now prove Lemma 12. According to Eq. 6, the worker $i \in Q_l$ with the least $\frac{c_i}{\rho_i}$ value will contribute the most amount of $t_i \rho_i$ in its region l . Thus, to prove that the constraint in Eq. 17 leads to $t_i^* \rho_i \leq \bar{t}p$ for all $i \in Q_l$, it is sufficient to show that the data contribution $t_l^M \rho_l^M$ from the participating worker i with the least $\frac{c_i}{\rho_i}$ value is less than or equal to the upper bound $\bar{t}p$, i.e., $t_l^M \rho_l^M \leq \bar{t}p$. Indeed, we substitute the $R_l^{\max}$ term from Eq. 17 into Eq. 16 to obtain $t_l^M \rho_l^M = \bar{t}p$. This implies that $t_i^* \rho_i \leq \bar{t}p$ for all $i \in Q_l$. Hence, the proof is complete. Note that the constraint on R_l simply constraints the feasible region and does not affect the concavity of the crowdsourcer's Problem 1. $\square$

Corollary The Stackelberg equilibrium can be shown to exist under the constraints in Theorem 10 but the uniqueness property of the unconstrained solution is lost. Furthermore, the solution can be obtained via the same algorithmic solution as utilized in the original unconstrained case, with a minor modification to Algorithm 1 as shown in Algorithm 2.

Finally, we have the following Theorem 14, which states the sufficient conditions where each worker's optimal data contribution satisfies the crowdsourcer's $\underline{t}p$ and $\bar{t}p$ bounds. This allows the crowdsourcer to estimate the number of participating workers in a region l given its chosen bounds $\underline{t}p$ and $\bar{t}p$, and reward allocation R_l .

Theorem Suppose Eq. 18 is true for all $i \in Q_l$. Then the worker i 's optimal data contribution in the Stackelberg equilibrium will satisfy the crowdsourcer's $\underline{t}p$ and $\bar{t}p$ bounds, i.e., $\underline{t}p \leq t_i^* \rho_i \leq \bar{t}p$:

$$\underline{t}p + \delta_i R_l \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right)^2 \leq \bar{t}p, \quad (18)$$

$$\text{where } \delta_i = \frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} \left(1 - \frac{\underline{t}p \sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{(|Q_{l_i}| - 1) R_l} \right) - \frac{c_i}{\rho_i} \text{ and } \delta_i \geq 0.$$

Proof We first show that Eq. 18 is a sufficient condition for $t_i^* \rho_i \leq \bar{t}p$. From Eq. 18, we obtain:

$$\begin{aligned}
 & \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} - \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} \right) + t\rho + \delta_i R_{l_i} \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right)^2 \leq \bar{t}\bar{p} \\
 & \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} - \left[\frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} - \left(\frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} \right)^2 \frac{t\rho}{R_{l_i}} - \delta_i \right] \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right)^2 R_{l_i} \leq \bar{t}\bar{p} \\
 & \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} - \left[\frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} \left(1 - \frac{t\rho}{(|Q_{l_i}| - 1)R_{l_i}} \sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j} \right) - \delta_i \right] \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right)^2 R_{l_i} \leq \bar{t}\bar{p} \\
 & \stackrel{(i)}{\Rightarrow} \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} - \left[\frac{c_i}{\rho_i R_{l_i}} \right] \left(\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} \right)^2 \leq \bar{t}\bar{p} \\
 & \stackrel{(ii)}{\Rightarrow} \sum_{j \in Q_{l_i}} t_j \rho_j - \frac{c_i}{\rho_i R_{l_i}} \left(\sum_{j \in Q_{l_i}} t_j \rho_j \right)^2 \leq \bar{t}\bar{p} \\
 & \stackrel{(iii)}{\Rightarrow} t_i^* \rho_i \leq \bar{t}\bar{p}, \\
 & \text{where (i) we substitute } \frac{c_i}{\rho_i} = \frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} \left(1 - \frac{t\rho}{(|Q_{l_i}| - 1)R_{l_i}} \sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j} \right) - \delta_i, \text{ (ii) we substitute } \sum_{j \in Q_{l_i}} t_j \rho_j = \\
 & \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} \text{ from Eq. 26, and (iii) we substitute } t_i^* \rho_i = \sum_{j \in Q_{l_i}} t_j \rho_j - \frac{c_i}{\rho_i R_{l_i}} \left(\sum_{j \in Q_{l_i}} t_j \rho_j \right)^2 \\
 & \text{from Eq. 25.}
 \end{aligned}$$

After proving that Eq. 18 is a sufficient condition for $t_i^* \rho_i \leq \bar{t}\bar{p}$, we show that Eq. 18 also leads to $t_i^* \rho_i \geq t\rho$. From Eq. 18, we have $\frac{c_i}{\rho_i} \leq \frac{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}}{|Q_{l_i}| - 1} (1 - \frac{t\rho}{(|Q_{l_i}| - 1)R_{l_i}} \sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j})$, which by Lemma 11 leads to $t_i^* \rho_i \geq t\rho$. Hence, the proof is complete. $\square$

Remarks: The feasible $t\rho$ and $\bar{t}\bar{p}$ for each worker i vary according to the its sensing cost c_i and location granularity ρ_i , which affect δ_i (the difference between the maximum allowable $\frac{c_i}{\rho_i}$ value in the set of participating workers and the worker i 's $\frac{c_i}{\rho_i}$ value). When $\delta_i = 0$, there is strict equality in Eq. 18 and $t\rho$ equals $\bar{t}\bar{p}$. However, as δ_i increases, we have $t\rho \ll \bar{t}\bar{p}$. For a fixed reward R_l , the feasible range between $t\rho$ and $\bar{t}\bar{p}$ increases as the number of participating workers Q_{l_i} increases.

4.2 Achieving Pareto Efficiency

We examine the Pareto efficiency (see Definition 15) of the Stackelberg equilibrium point in our proposed Stackelberg incentive model and study how to achieve efficiency. At the Pareto efficient equilibrium point, the crowdsourcer's rewards are allocated in such a way

that it is not possible to reallocate them to increase the utility of any individual (including the crowdsourcer itself) without making at least one individual's utility decrease. In other words, it implies an efficient allocation of resources.

Definition Pareto Efficiency A strategy profile $(\mathbf{R}^P, \mathbf{t}^P)$ is Pareto efficient if there exists no other strategy $(\mathbf{R}, \mathbf{t})$ where $\mathbf{R} \geq 0, \mathbf{t} \geq 0$ such that:

$$U_{CS}(\mathbf{R}; \mathbf{t}) \geq U_{CS}(\mathbf{R}^P; \mathbf{t}^P),$$

$$u_i(t_i; \mathbf{t}_{-i}, \mathbf{R}) \geq u_i(t_i^P; \mathbf{t}_{-i}^P, \mathbf{R}^P), \quad \forall i \in \mathcal{I},$$

with at least one strict inequality.

Theorem The proposed (unbounded) Stackelberg game has a unique Stackelberg equilibrium $(\mathbf{R}^{SE}, \mathbf{t}^{SE})$ that may not be Pareto efficient.

Proof: See Appendix 6.

By Theorem 16, the proposed (unbounded) Stackelberg game has a unique Stackelberg equilibrium $(\mathbf{R}^{SE}, \mathbf{t}^{SE})$ that may not be Pareto efficient. To achieve efficiency of the Stackelberg equilibrium, we first define a social welfare function $w(\mathbf{R}, \mathbf{t})$ to be the weighted sum of the crowdsourcer and the workers' utilities:

$$w(\mathbf{R}, \mathbf{t}) = \gamma_{CS} U_{CS}(\mathbf{R}; \mathbf{t}) + \sum_{i \in \mathcal{I}} \gamma_i u_i(t_i; \mathbf{t}_{-i}, R_{l_i})$$

$$= \gamma_{CS} \sum_{i \in \mathcal{I}} \alpha_i t_i^\beta + \sum_{i \in \mathcal{I}} \gamma_i \left(\frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{l_i} - c_i t_i \right), \quad (19)$$

for some weights $\gamma_{CS}, \gamma_1, \dots > 0$.

It is well-known that any allocation which maximizes a social welfare function is also Pareto efficient. Thus, to achieve efficiency of the Stackelberg equilibrium, we can introduce the penalty function:

$$\Psi(\mathbf{t}) = \sum_{i \in \mathcal{I}} \gamma_i \left(\frac{t_i \rho_i}{\sum_{j \in \mathcal{Q}_{l_i}} t_j \rho_j} R_{l_i} - c_i t_i \right) \text{ to the crowdsourcer's utility function, i.e.,}$$

$$U_{CS}(\mathbf{R}; \mathbf{t}) = \gamma_{CS} \sum_{i \in \mathcal{I}} \alpha_i t_i^\beta + \Psi(\mathbf{t}). \quad (20)$$

One way to encourage the crowdsourcer to maximize $w(\mathbf{R}, \mathbf{t})$ would be to introduce a third party regulator that can offer tax rebates proportional to the weighted sum of the worker's utilities $\Psi(\mathbf{t})$, thus contributing to the crowdsourcer's utility. Note that the penalty function $\Psi(\mathbf{t})$ is a convex function of $\mathbf{R}$ and hence does not affect the convexity of the crowdsourcer's Problem 1. Therefore, the unique Stackelberg equilibrium solution still exists.

5 Simulation Study

To evaluate the performance of the proposed Stackelberg incentive model in a real-world spatial data sensing application, we design a spatial data sensing case study application where we assume that the crowdsourcer wishes to perform a spatial estimation task.

To achieve this, the crowdsourcer requires data from mobile smartphone users (workers) located in a range of different locations. The greater the spatial diversity obtained in the data, in general the easier it will be for the crowdsourcer to perform the spatial estimation task. In particular, we will design the case study to undertake a challenge of spatial prediction. The dataset collected by the crowdsourcer will be obtained by application of our spatial privacy-preserving data sharing mechanism based on our proposed Stackelberg incentive model (Algorithm 2), which attempts to maximize spatial coverage.

We consider a real-world mobile crowd sensing problem such as the spatial monitoring and prediction of environmental temperature Chen et al. (2015) and Mun et al. (2009). This involves the crowdsourcer incentivizing and paying the workers for the data they collected from their spatially placed sensors. Using the collected dataset, the crowdsourcer wishes to make a spatial estimation, achieved by conducting a spatial regression based on a Gaussian process Rasmussen and Williams (2005), estimated from the collected data. We then evaluate the spatial regression performance of the proposed Stackelberg incentive model against two baseline non-game-theoretic incentive schemes that seek to maximize the spatial coverage of the collected dataset.

5.1 Baseline Coverage Metrics

We consider the two baseline coverage maximizing schemes: (i) the location-based incentive mechanism proposed in Jaimes et al. (2012), which maximizes a geometric disk coverage model, and (ii) the work in Xiong et al. (2016), which maximize a k -depth coverage model. The two coverage models (see Fig. 3) are detailed as follows.

5.1.1 Geometric Disk Coverage Metric

The geometric disk coverage scheme was proposed in Jaimes et al. (2012) to measure the coverage $c(\mathbf{x}_i)$ of a sensor data from a precise (uncloaked) location $\mathbf{x}_i$:

$$c(\mathbf{x}_i) = \begin{cases} 1 & \text{if } \|\mathbf{x}_i - \mathbf{x}_j\|_2 \leq r, \\ 0 & \text{otherwise,} \end{cases}$$

(i) Geometric Disk

(ii) k -depth

Fig. 3 Illustration of the two coverage metrics used

where $\|\cdot\|_2$ denotes the Euclidean distance and r is the sensed radius of the sensor data.

To optimize the disk coverage metric, the crowdsourcer greedily buys the minimum amount of data $\underline{t}$ from each worker in regions where $c(\mathbf{x}_i) = 0$, starting with the cheapest worker first. For a fair comparison with our proposed Stackelberg incentive model, we let $c(\mathbf{x}_i) = 1$ when $\mathbf{x}_i \in l$ where l is a partitioned region.

5.1.2 k -depth Coverage Metric

The following k -depth coverage model (and its variants) was proposed in Xiong et al. (2016) to measure coverage of a set of N sensor data $t_1, \dots, t_N$ from a region l where the coverage $c(t_1, \dots, t_N) = \min(N, k)$ or equivalently:

$$c(t_1, \dots, t_N) = \begin{cases} N & \text{if } N \leq k, \\ k & \text{otherwise,} \end{cases}$$

where k is the depth parameter.

To optimize the k -depth coverage metric, the crowdsourcer greedily buys the minimum amount of data $\underline{t}$ from k workers in each region where $c(t_i) = 0$, starting with the cheapest worker first.

5.2 Simulation Setup

We use the temperature measurements from the Intel lab dataset Madden (2004), which contains the temperature, humidity, light, voltage, connectivity, and location information collected from 54 sensor nodes deployed in the Intel Berkeley Research lab between February 28th and April 5th, 2004. We partitioned the lab's spatial area into the eight regions l as shown in Fig. 4. The proposed Stackelberg incentive model and the baseline coverage schemes are then used to purchase a subset of the available temperature data. We took a one-hour interval (from 01:00–02:00, 28/2/2004) from the dataset and apply the Gaussian process regression technique Rasmussen and Williams (2005) (a supervised learning technique for regression) to evaluate how the two coverage metrics correspond to the actual amount of predictive uncertainty. The (Gaussian) radial basis function (RBF) kernel Rasmussen and Williams (2005) is used for our Gaussian process regression.

We briefly introduce the main idea behind the Gaussian process regression. Given a set of n input training location vector $\mathbf{x}$ and observations $\mathbf{y}$, we assume that the observed $\mathbf{y}$ are

Fig. 4 Partitioned regions of the Intel lab, which contains 54 sensors nodes

generated by some latent function f plus an independent and identically distributed Gaussian noise with zero mean and variance σ_y^2 . Suppose there are n_* test points where we are interested in obtaining the predicted observation values. Let $k(\cdot, \cdot)$ be a covariance function and let $K(X, X_*)$ denote the $n \times n_*$ (kernel) matrix of the covariances evaluated at all pairs of training and test points, and similarly for $K(X, X)$, $K(X_*, X_*)$ and $K(X_*, X)$. The predictive (posterior) mean $\tilde{f}_*(x_*)$ and variance $V(x_*)$ for the Gaussian process regression for a new input test vector x_* are given by:

$$\begin{aligned}\tilde{f}_*(x_*) &= K(X_*, X)[K(X, X) + \sigma_y^2 I]^{-1} \mathbf{y}, \\ V(x_*) &= K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_y^2 I]^{-1} K(X, X_*).\end{aligned}\quad (21)$$

There are a number of publicly available libraries that implement the Gaussian process regression and we chose to use Python's scikit-learn machine learning library Pedregosa et al. (2011) in our simulations.

5.2.1 Test Scenarios

The following two test scenarios were examined.

Scenario (I): Spatial regression of the two cross intersections of interest (second and fourth column from the left side of Fig. 4) where no sensor data is available. We assume that each worker i 's sensing cost c_i is inversely proportional to its distance from the nearest cross intersection of interest, i.e., workers have a higher sensing cost if they are located near the two cross intersections and their location granularities ρ_i are inversely proportional to the number of workers in their respective regions, i.e., workers located in denser regions are less privacy-sensitive. We use the predictive variance $V(x_*)$ of the spatial Gaussian process regression in the two cross intersections (points of interest) as the metric for comparison. Intuitively, a lower predictive variance implies better predictive performance.

Scenario (II): Spatial regression of the entire spatial area shown in Fig. 4. We assume that the workers' sensing costs c_i are uniformly selected from $[0.25, 0.5, 0.75, 1]$ and their location granularities ρ_i are inversely proportional to the number of workers in their respective regions. We conducted spatial Gaussian process regression to obtain the predicted (mean) temperature $\tilde{f}_*(x_*)$ of all the sensor locations. We use the mean squared error (MSE) values (computed by taking the difference between the predicted temperature and the actual temperature measurement from the dataset) as the metric for comparison. Intuitively, a lower MSE value implies better predictive performance.

Table 2 Predictive standard deviation values for Scenario (I).

Scheme	Values	
	Point 1	Point 2
Disk	0.3435	0.3597
k -depth	0.2707	0.2620
Proposed	0.2707	0.2620

Table 3 Baseline coverage scores for Scenario (I).

Score Scheme	Disk	k -depth ($k = 2$)
Disk	8	8
k -depth	8	16
Proposed	8	16

5.2.2 Simulation Parameters

We set the sensing cost $c_i \in (0, 1]$, location granularity $\rho_i \in [1, 2]$, budget $R^{\text{budget}} = 10$, $R_l^{\min} = 0$ for all $l \in \mathcal{L}$, $R_l^{\max} = \infty$ for all $l \in \mathcal{L}$, minimum data $\underline{t} = 1$, and maximum data $\bar{t} = 3$. We chose a budget constraint R^{budget} to limit the number of purchased data, i.e., the scheme cannot afford to purchase all the available temperature data. In addition, the amount of budget spent by all three schemes are capped at the same amount. In the two baseline coverage schemes, we offer each worker their sensing cost for each unit of data. We let the crowdsourcer's system parameter $\alpha_i = \rho_i$ and $\beta = 0.1$ in our proposed Stackelberg incentive model. We varied the k value of the baseline k -depth coverage scheme.

5.3 Simulation Results and Discussion

We now discuss the simulation results for the test scenarios (I) and (II).

Scenario (I): We list the predictive variances of the two baseline coverage schemes and the proposed Stackelberg incentive model at the two cross intersections of interest in Table 2, and the corresponding coverage scores in Table 3.

From the results, we observe that the proposed Stackelberg incentive model has better predictive variances in the two cross intersections of interest compared to the baseline disk coverage scheme. The predictive variances from the k -depth coverage scheme and the proposed Stackelberg incentive model are the same for all values of $k > 1$. This is because the sensing costs of the workers are the lowest when the workers are located far away from the two cross intersections of interest and under the limited budget constraint, the k -depth coverage scheme and the proposed Stackelberg incentive model will selected the same set of participating workers. Note that the used disk coverage scheme is equivalent to the k -depth coverage scheme when $k = 1$. This is due to the usage of the partitioned regions instead of

Fig. 5 Heatmap of the predictive standard deviation for **i** disk, and **ii** the k -depth and proposed Stackelberg incentive models - Scenario (I)

Table 4 Mean squared error values for Scenario (II).

Scheme \ Values	MSE	Standard Deviation
Disk	35.0554	0.5228
k -depth ($k = 2$)	26.5106	0.3338
k -depth ($k = 3$)	26.3102	0.2832
k -depth ($k = 4$)	32.9805	0.3034
k -depth ($k = 5$)	39.4518	0.3431
Proposed	25.5666	0.2856

the original geometric disk coverage areas, which are not applicable to our scenario. To visualize the locations of the participating workers and the predictive variances, we plot the heat map of the predictive variances for the baseline coverage schemes and the proposed Stackelberg incentive model in Fig. 5. The two cross labels in the figures represent the two points of interest where no sensor data is available. As the disk coverage scheme only selects one worker per region, it does not collect enough data points compared to the other two schemes.

Scenario (II): We list the MSE values of the two baseline coverage schemes and the proposed Stackelberg incentive model for the entire spatial area of interest in Table 4, and the corresponding coverage scores in Table 5. From the results, we observe that the proposed Stackelberg incentive model has better predictive performance (lower MSE value) in the entire spatial area of interest compared to the two baseline coverage schemes. While the baseline coverage schemes simply prioritize data from cheaper workers, our Stackelberg incentive model is able to offset the higher sensing costs of the workers with the α_i parameters. To visualize the location of the participating workers and the predicted mean temperature values, we plot the heat map of the predicted mean values for the baseline disc and k -depth schemes for $k = 3$ (where the MSE value is the lowest) and the proposed Stackelberg incentive model in Fig. 6.

5.3.1 Accommodating Location Uncertainty

The Gaussian process regression technique accommodates location uncertainty of the workers' sensing data due to the use of cloaking regions. Assuming that the input locations of the sensing data $\mathbf{x}$ are corrupted by i.i.d. Gaussian noise with the noise variance set to the square of the approximated radius of the cloaking regions, the Gaussian process regression model proposed in Mchutchon and Rasmussen (2011) can be applied to account for location uncertainty. Mainly, an additional corrective term proportional to the gradient of the posterior mean can be added into the noise term in Eq. 21 to account for the location uncertainty of the training inputs.

Table 5 Baseline coverage scores for Scenario (II).

Scheme \ Score	Disk	k -depth ($k = 2$)	k -depth ($k = 3$)	k -depth ($k = 4$)	k -depth ($k = 5$)
Disk	8	8	8	8	8
k -depth	8	16	22	23	23
Proposed	6	12	18	22	23

Fig. 6 Heatmap of the predictive means in Scenario II. Left: disk; Middle: k -depth; Right: proposed Stackelberg incentive model

6 Conclusion

We designed a privacy-aware Stackelberg incentive mechanism that improves the spatial coverage of the collected dataset. Our proposed incentive model is privacy-aware, in that it allows privacy-sensitive mobile smartphone users to submit coarse-grained (or quantized) location information to the crowdsourcer. We studied the properties of the proposed Stackelberg incentive model analytically and presented efficient algorithmic solutions. We also extended the basic model to accommodate bounds on the users' data contributions and studied how Pareto efficiency can be achieved. We showed via simulations using a real-world sensing dataset that our proposed incentive model produced greater spatial diversity in sourced data, leading to better model predictive performance compared to two other coverage-maximizing schemes that maximize a different coverage metric. For future work, it would be interesting to extend our (static game) Stackelberg model for dynamic games played over a period of time where the smartphone users are allowed to move between regions. While our Stackelberg equilibrium is stable in the studied static model played by the crowdsourcer and the mobile smartphone users in one time period, a different notation of equilibrium needs to be considered for the dynamic setting.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix 1. Proof of Lemma 2

A unique Nash equilibrium exists in the Followers game if for all $i \in \mathcal{I}$ Rosen (1965): (i) the domain of the workers' strategy set $\{t_i\}_{i \in \mathcal{I}}$ is convex and compact, and (ii) u_i is continuous and strictly concave in t_i . Indeed, the domain of the workers' strategy set $\{t_i\}_{i \in \mathcal{I}}$ is convex and compact since t_i is assumed to be bounded, and u_i is continuous in t_i , and strictly concave in t_i as $\frac{\partial^2 u_i}{\partial t_i^2} < 0$:

$$\frac{\partial u_i}{\partial t_i} = \frac{\rho_i \sum_{j \in Q_{l_i}: j \neq i} t_j \rho_j}{\left(\sum_{j \in Q_{l_i}} t_j \rho_j \right)^2} R_{l_i} - c_i,$$

$$\frac{\partial^2 u_i}{\partial t_i^2} = \frac{-2\rho_i^2 \sum_{j \in Q_{l_i}: j \neq i} t_j \rho_j}{\left(\sum_{j \in Q_{l_i}} t_j \rho_j \right)^3} R_{l_i} < 0. \quad (22)$$

Therefore, a Nash equilibrium exists in the Followers game. $\square$

Appendix 2. Proof of Theorem 3

By Lemma 2, there exists a unique strategy profile that maximizes the utility of each worker given the strategies of the other workers. Thus, if each worker i plays its best response strategy, it will achieve the unique Nash equilibrium point t_i^* . To prove Theorem 3, we derive t_i^* by solving $\frac{\partial u_i}{\partial t_i} = 0$ to obtain:

$$\rho_i R_{l_i} \sum_{j \in Q_{l_i}: j \neq i} t_j^* \rho_j = c_i \left(\sum_{j \in Q_{l_i}} t_j^* \rho_j \right)^2. \quad (23)$$

Next, we manipulate the expression in Eq. 23 to obtain:

$$\left(\sum_{j \in Q_{l_i}} t_j^* \rho_j - t_i^* \rho_i \right) \rho_i R_{l_i} = c_i \left(\sum_{j \in Q_{l_i}} t_j^* \rho_j \right)^2 \quad (24)$$

$$\stackrel{(i)}{\Rightarrow} t_i^* \rho_i = \left(\sum_{j \in Q_{l_i}} t_j^* \rho_j \right) \left[1 - \frac{c_i}{\rho_i R_{l_i}} \left(\sum_{j \in Q_{l_i}} t_j^* \rho_j \right) \right] \quad (25)$$

$$\stackrel{(ii)}{\Rightarrow} \sum_{j \in Q_{l_i}} t_j^* \rho_j = |Q_{l_i}| \sum_{j \in Q_{l_i}} t_j^* \rho_j - \frac{1}{R_{l_i}} \left(\sum_{j \in Q_{l_i}} t_j^* \rho_j \right)^2 \sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j} \quad (26)$$

$$\stackrel{(iii)}{\Rightarrow} \sum_{j \in Q_{l_i}} t_j^* \rho_j = \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i},$$

where (i) we express $t_i^* \rho_i$ in terms of the other $t_j^* \rho_j$ values, (ii) we sum up the $t_i^* \rho_i$ values in Eq. 25 for all participating workers in region l_i , and (iii) we divide the previous expression by $\sum_{j \in Q_{l_i}} t_j^* \rho_j$.

Finally, we substitute Eq. 26 into Eq. 25 to obtain the unique Nash equilibrium point for each worker i as required. $\square$

Appendix 3. Proof of Lemma 4

To prove Lemma 4, we show that the set of participating workers $\mathcal{Q}_l$ computed by Algorithm 1 always satisfies Eq. 7 for all regions $l \in \mathcal{L}$.

Algorithm 1 Compute the Nash equilibrium solution of the Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$.

```

1 function SolveFollowersGame( $\mathbf{c}, \rho, \mathbf{l}, \mathbf{R}$ ):
   Input : sensing costs  $\mathbf{c}_{1,\dots,N}$ , location granularities  $\rho_{1,\dots,N}$ , workers' regions  $\mathbf{l}_{1,\dots,N}$ ,
           rewards  $\mathbf{R}_{1,\dots,L}$ .
   Output: data sold to crowdsourcer  $\mathbf{t}^*_{1,\dots,N}$ .
2 foreach region  $l \in \mathcal{L}$  do
3   Sort workers in region  $l$  according to their privacy-weighted cost  $\frac{c_i}{\rho_i}$  in ascending
   order where  $\frac{c_i}{\rho_i} \leq \frac{c_{i+1}}{\rho_{i+1}}$ .
4   Let  $\mathcal{Q}_l = \{1, 2\}$  be the set of participating workers with  $t_i^* > 0$ .
5   Set  $\mathcal{Q}_l \leftarrow \mathcal{Q}_l \cup \{i\}$  for each worker  $i = 3, \dots$  in region  $l$  if the condition in Eq. 7
   is met.
   (Note: the looping can stop at the  $i$ th step when the condition is not met.)
6   Set  $t_i^*$  according to Eq. 6 for all workers in region  $l$ .
7 end
```

Algorithm 2 Compute the bounded Nash equilibrium solution for the Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$.

```

1 function SolveBoundedFollowerGame( $\mathbf{c}, \rho, \mathbf{l}, \mathbf{R}, \underline{t\rho}, \overline{t\rho}$ ):
   Input : sensing costs  $\mathbf{c}_{1,\dots,N}$ , location granularities  $\rho_{1,\dots,N}$ , workers' regions  $\mathbf{l}_{1,\dots,N}$ ,
           rewards  $\mathbf{R}_{1,\dots,L}$ .
   Output: data sold to crowdsourcer  $\mathbf{t}^*_{1,\dots,N}$ .
2 foreach region  $l \in \mathcal{L}$  do
3   Compute unbounded  $t_i$  for all workers in region  $l$  using Algorithm 1.
4   if there exists more than one participating worker  $i$  with  $t_i \rho_i \geq \underline{t\rho}$  then
5     Sort workers in region  $l$  according to their privacy-weighted cost  $\frac{c_i}{\rho_i}$  in
     ascending order where  $\frac{c_i}{\rho_i} \leq \frac{c_{i+1}}{\rho_{i+1}}$ .
6     Let  $\mathcal{Q}_l = \{1, 2\}$  be the set of participating workers with  $t_i^* > 0$ .
7     Set  $\mathcal{Q}_l \leftarrow \mathcal{Q}_l \cup \{i\}$  for each worker  $i$  in region  $l$  if condition Eq. 13 of
     Theorem 10 is met. (note: the looping can stop at the  $i$ th step when the
     condition is not met.)
8     Set  $t_i^*$  according to Eq. 6 for all workers in region  $l$ .
9   else
10    Set  $t_i^* = 0$  for all workers  $i$  in region  $l$ .
11 end
```

Consider the region I . Suppose the constraint in Eq. 7 is not satisfied, then the following constraints must be true:

$$\begin{aligned}\frac{c_i}{\rho_i} &\geq \frac{\sum_{j \in Q_I: j \neq i} \frac{c_j}{\rho_j} + \frac{c_i}{\rho_i}}{|Q_I| - 1} \\ \frac{c_i}{\rho_i} &\geq \frac{\sum_{j \in Q_I: j \neq i} \frac{c_j}{\rho_j}}{|Q_I| - 2}.\end{aligned}\quad (27)$$

From Eq. 27, it is obvious that if $\frac{c_k}{\rho_k} \geq \frac{c_i}{\rho_i}$ and $\frac{c_i}{\rho_i} \geq \frac{\sum_{j \in Q_I: j \neq i} \frac{c_j}{\rho_j}}{|Q_I| - 2}$, then $\frac{c_k}{\rho_k} \geq \frac{\sum_{j \in Q_I: j \neq i} \frac{c_j}{\rho_j}}{|Q_I| - 2}$. Hence, the greedy computation of the set of participating workers Q_I from the sorted $\frac{c_i}{\rho_i}$ values in Algorithm 1 does not affect the set of participating workers Q_I that achieves the unique Nash equilibrium solution of the Followers game. $\square$

Appendix 4. Proof of Theorem 5

To show that the obtained solution from Algorithm 1 is the unique Nash equilibrium (NE) solution of the workers, we proof that the $\frac{\partial u_i}{\partial t_i} = 0$ (stationary point) condition given in Eq. 24 is satisfied by the set of participating workers $i \in Q$. By Lemma 2, there exists a unique NE in the followers game. Hence, any stationary point is the unique NE point for the

workers. From Algorithm 1, we have $t_i^* = \frac{|Q_I| - 1}{\sum_{j \in Q_I} \frac{c_j}{\rho_j}} \left(1 - \frac{(|Q_I| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_I} \frac{c_j}{\rho_j}} \right) \frac{R_{l_i}}{\rho_i}$ if $i \in Q_{l_i}$, and $t_i^* = 0$ otherwise. In addition, we have the expression $\sum_{j \in Q_{l_i}} t_j^* \rho_j = \frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i}$ from Eq. 26.

We substitute the expressions for t_i^* and $\sum_{j \in Q_{l_i}} t_j^* \rho_j$ into the $\frac{\partial u_i}{\partial t_i}$ expression in Eq. 24 to obtain the following equality:

$$\frac{(|Q_{l_i}| - 1) R_{l_i}}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \left(\frac{(|Q_{l_i}| - 1) \frac{c_i}{\rho_i}}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} \right) R_{l_i} = \frac{c_i}{\rho_i} \left[\frac{|Q_{l_i}| - 1}{\sum_{j \in Q_{l_i}} \frac{c_j}{\rho_j}} R_{l_i} \right]^2. \quad (28)$$

Since Eq. 28 satisfies the expression for $\frac{\partial u_i}{\partial t_i} = 0$, we conclude that Algorithm 1 correctly outputs the unique Nash equilibrium solution of the workers in Q_I . Also, by Lemma 4, Algorithm 1 also correctly computes the set of participating workers Q_I as used in Eq. 28. $\square$

Appendix 5. Proof of Lemma 8

From Eq. 11, we have the following expression for all workers $i \in \mathcal{Q}_l$:

$$\begin{aligned} \frac{\partial u_i}{\partial c'_i} &= -(|\mathcal{Q}_l| - 1) \frac{\left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2 + \frac{c'_i}{\rho_i} \left((|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right) - |\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}}{\left(\sum_{j \in \mathcal{Q}_l} \frac{c'_j}{\rho_j} \right)^3} R_{l_i}, \\ \frac{\partial^2 u_i}{\partial (c'_i)^2} &= 2(|\mathcal{Q}_l| - 1) \\ &\quad \times \frac{\left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2 + \frac{c'_i}{\rho_i} \left((|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right) - (2|\mathcal{Q}_l| - 1) \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}}{\left(\sum_{j \in \mathcal{Q}_l} \frac{c'_j}{\rho_j} \right)^4} R_{l_i}. \end{aligned} \quad (29)$$

From Eq. 29, we set $\frac{\partial u_i}{\partial c'_i} = 0$ to obtain the critical point:

$$\begin{aligned} |\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} &= \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2 + \frac{c'_i}{\rho_i} \left((|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right), \\ \frac{c'_i}{\rho_i} &= \frac{|\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} - \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2}{(|\mathcal{Q}_l| - 2) \left(\frac{c_i}{\rho_i} \right) + \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}}. \end{aligned} \quad (30)$$

We substitute the expression of the critical point derived in Eq. 30 into the $\frac{\partial^2 u_i}{\partial (c'_i)^2}$ expression in Eq. 29 to obtain:

$$\begin{aligned} \frac{\partial^2 u_i}{\partial (c'_i)^2} &\propto \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2 + |\mathcal{Q}_l| \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} - \left(\sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} \right)^2 - (2|\mathcal{Q}_l| - 1) \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j}, \\ &\propto \left(\frac{c_i}{\rho_i} \right) \sum_{j \in \mathcal{Q}_l: j \neq i} \frac{c'_j}{\rho_j} (|\mathcal{Q}_l| - (2|\mathcal{Q}_l| - 1)), \\ &\propto 1 - |\mathcal{Q}_l|. \end{aligned} \quad (31)$$

The denominator term $\sum_{j \in \mathcal{Q}_l} \frac{c'_j}{\rho_j}$ in the $\frac{\partial^2 u_i}{\partial (c'_i)^2}$ expression from Eq. 29 is positive due to

both c_i and ρ_i being positive. Since we assume that $|\mathcal{Q}_l| > 2$, then we have $\frac{\partial^2 u_i}{\partial (c'_i)^2} < 0$. This leads us to conclude that the critical point is a maximum point. Hence, the proof is complete. $\square$

Appendix 6. Proof of Theorem 16

To prove Theorem 16, we make the following two claims.

Claim Suppose that the α_i , ρ_i , and c_i values are constant for all workers $i \in \mathcal{I}$, the Stackelberg equilibrium of the proposed Stackelberg incentive model is Pareto efficient.

Proof To prove the claim, we use the following key observations:

$$\sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j; \mathbf{t}_{-j}, R_{l_j}) = R_{l_i} - \sum_{j \in \mathcal{Q}_{l_i}} c t_j. \quad (32)$$

$$U_{CS}(\mathbf{R}; \mathbf{t}) = \alpha \sum_{i \in \mathcal{I}} t_i^\beta. \quad (33)$$

Consider a strategy profile $(\mathbf{R}, \mathbf{t}) \neq (\mathbf{R}^{SE}, \mathbf{t}^{SE})$. If $U_{CS}(\mathbf{R}; \mathbf{t}) > U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$, then $\exists i$ where $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) < u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$. It can be shown that the previous statement is true since $\sum_{j \in \mathcal{Q}_{l_i}} t_j > \sum_{j \in \mathcal{Q}_{l_i}} t_j^{SE}$ is a necessary condition for $U_{CS}(\mathbf{R}; \mathbf{t}) > U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$

when the α_i , ρ_i , and c_i values are constant for all workers $i \in \mathcal{I}$. From Eq. 32, we know that $\sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j; \mathbf{t}_{-j}, R_{l_j})$ is inversely proportional to $\sum_{j \in \mathcal{Q}_{l_i}} c t_j$. This means that if $\sum_{j \in \mathcal{Q}_{l_i}} t_j > \sum_{j \in \mathcal{Q}_{l_i}} t_j^{SE}$ is true, then we have $\sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j; \mathbf{t}_{-j}, R_{l_j}) < \sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j^{SE}; \mathbf{t}_{-j}^{SE}, R_{l_j}^{SE})$. Hence, we conclude that $\exists i$ where $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) < u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$.

Similarly, it can be shown that if $\exists i$ where $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) > u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$ and $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, $\forall i \in \mathcal{I}$, then $U_{CS}(\mathbf{R}; \mathbf{t}) < U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$. This is because if $\exists i$ where $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) > u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$ and $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, $\forall i \in \mathcal{I}$, then we have $\sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j; \mathbf{t}_{-j}, R_{l_j}) > \sum_{j \in \mathcal{Q}_{l_i}} u_j(t_j^{SE}; \mathbf{t}_{-j}^{SE}, R_{l_j}^{SE})$. This is only possible if $\sum_{j \in \mathcal{Q}_{l_i}} c t_j < \sum_{j \in \mathcal{Q}_{l_i}} c t_j^{SE}$, which means $\sum_{j \in \mathcal{Q}_{l_i}} t_j < \sum_{j \in \mathcal{Q}_{l_i}} t_j^{SE}$. Therefore, from Eq. 33, we conclude that $U_{CS}(\mathbf{R}; \mathbf{t}) < U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$.

Hence, the proof is complete. $\square$

Claim The proposed Stackelberg incentive model may not have a Pareto efficient Stackelberg equilibrium. In other words, suppose that $(\mathbf{R}, \mathbf{t}) \neq (\mathbf{R}^{SE}, \mathbf{t}^{SE})$ and $U_{CS}(\mathbf{R}; \mathbf{t}) > U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$, we have $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, $\forall i \in \mathcal{I}$.

Proof Consider the scenario where there are 2 regions l_1 and l_3 with 2 workers each. Let $\rho_i = 1$, $c_i = 1$ for the workers in the 2 regions. Let $l_1 = l_2$ and $l_3 = l_4$, $\alpha_1 = \alpha_2$, $\alpha_3 = \alpha_4$, and $\alpha_1 < \alpha_3$. Intuitively, given that the worker costs are the same but $\alpha_1 < \alpha_3$, then $R_{l_1}^{SE} < R_{l_3}^{SE}$.

Suppose we have $R_{l_3} > R_{l_3}^{SE}$ and $R_{l_1} < R_{l_1}^{SE}$ (recall that the total rewards is bounded, so R_{l_1} must decrease if R_{l_3} increases). Let $R_{l_3} = R_{l_3}^{SE} + \Delta$ and $R_{l_1} = R_{l_1}^{SE} - \Delta$ where $\Delta > 0$.

First, we obtain the closed-form expression for t_i at the Stackelberg equilibrium. We substitute the ρ_i and c_i values into Eq. 6 and obtain:

$$t_i^{SE} = \frac{R_{l_i}^{SE}}{2} \left(1 - \frac{1}{2}\right) = \frac{R_{l_i}^{SE}}{4}, \forall i \in \mathcal{I}. \quad (34)$$

Next, we substitute the t_i expression in Eq. 34 into Eq. 3 to obtain

$$u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE}) = \frac{R_{l_i}^{SE}}{2} - \frac{R_{l_i}^{SE}}{4} = \frac{R_{l_i}^{SE}}{4}. \quad (35)$$

Suppose $u_1(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_1(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, from Eq. 3, we have:

$$\begin{aligned} \frac{R_{l_1}^{SE} - \Delta}{2} - t_1 &\geq \frac{R_{l_1}^{SE}}{4}, \\ t_1 &\leq \frac{R_{l_1}^{SE} - 2\Delta}{4}, \\ t_1 &= \frac{R_{l_1}^{SE} - 2\Delta}{4} - \Delta', \end{aligned} \quad (36)$$

where $\Delta' \geq 0$.

Similarly, if $u_3(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_3(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, from Eq. 3, we have:

$$\begin{aligned} \frac{R_{l_3}^{SE} + \Delta}{2} - t_3 &\geq \frac{R_{l_3}^{SE}}{4}, \\ t_3 &\leq \frac{R_{l_3}^{SE} + 2\Delta}{4}, \\ t_3 &= \frac{R_{l_3}^{SE} + 2\Delta}{4} - \Delta', \end{aligned} \quad (37)$$

where $\Delta' \geq 0$.

If $U_{CS}(\mathbf{R}; \mathbf{t}) > U_{CS}(\mathbf{R}^{SE}; \mathbf{t}^{SE})$, and $u_i(t_i; \mathbf{t}_{-i}, R_{l_i}) \geq u_i(t_i^{SE}; \mathbf{t}_{-i}^{SE}, R_{l_i}^{SE})$, $\forall i \in \mathcal{I}$ is true, then the following condition on U_{CS} must be true. From Eq. 8, we have:

$$\begin{aligned} \alpha_1 t_1^\beta + \alpha_2 t_2^\beta + \alpha_3 t_3^\beta + \alpha_4 t_4^\beta &> \alpha_1 (t_1^{SE})^\beta + \alpha_2 (t_2^{SE})^\beta + \alpha_3 (t_3^{SE})^\beta + \alpha_4 (t_4^{SE})^\beta, \\ \alpha_1 \left(\frac{R_{l_1}^{SE} - 2\Delta}{4} - \Delta' \right)^\beta + \alpha_3 \left(\frac{R_{l_3}^{SE} + 2\Delta}{4} - \Delta' \right)^\beta &> \alpha_1 \left(\frac{R_{l_1}^{SE}}{4} \right)^\beta + \alpha_3 \left(\frac{R_{l_3}^{SE}}{4} \right)^\beta. \end{aligned} \quad (38)$$

□

Therefore, Claim 2 shows a scenario where the proposed Stackelberg equilibrium may not be Pareto efficient for the crowdsourcer.

References

- Andrés ME, Bordenabe NE, Chatzikokolakis K, Palamidessi C (2013) Geo-indistinguishability: Differential privacy for location-based systems. In: Proc ACM Conference on Computer and Communications Security (CCS), pp 901–914
- Chen PY, Cheng SM, Ting PS, Lien CW, Chu FJ (2015) When crowdsourcing meets mobile sensing: a social network perspective. IEEE Commun. Mag 53(10):157–163. <https://doi.org/10.1109/MCOM.2015.7295478>. ISSN 0163-6804

- Cheng R, Zhang Y, Bertino E, Prabhakar S (2006) Preserving user location privacy in mobile data management infrastructures. In: Proc. Int. Workshop on Privacy Enhancing Technologies (PET), pages 393–412, Jun.
- Feng Z, Zhu Y, Zhang Q, Ni LM, Vasilakos AV (2014) TRAC: truthful auction for location-aware collaborative sensing in mobile crowdsourcing. In: Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM), pages 1231–1239, Apr. <https://doi.org/10.1109/INFOCOM.2014.6848055>
- Guo B, Wang Z, Yu Z, Wang Y, Yen NY, Huang R, Zhou X (2015) Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm. *ACM Comput. Surv.* 48(1):7–31. ISSN 0360-0300
- Jaimes LG, Vergara-Laurens I, Labrador MA (2012) A location-based incentive mechanism for participatory sensing systems with budget constraints. In: Proc. IEEE Int. Conf. Pervasive Comput. and Commun., pages 103–108, March. <https://doi.org/10.1109/PerCom.2012.6199855>
- Kido H, Yanagisawa Y, Satoh T (2005) An anonymous communication technique using dummies for location-based services. In: Proc. Int. Conf. Pervasive Services., pages 88–97, Jul.
- Koh JY, Peters GW, Leong D, Nevat I, Wong W-C (2017) Privacy-aware incentive mechanism for mobile crowd sensing. In: Proc. IEEE Int. Conf. Commun. (ICC)
- Lin J, Liu B, Sadeh N, Hong JI (2014) Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings. In: Proc. Symp. Usable Privacy and Security (SOUPS)
- Luo S, Sun Y, Ji Y, Zhao D (2016a) Stackelberg game based incentive mechanisms for multiple collaborative tasks in mobile crowdsourcing. *Mobile Networks and Applications* 21(3):506–522. <https://doi.org/10.1007/s11036-015-0659-3>. ISSN 1572-8153
- Luo T, Das SK, Tan HP, Xia L (2016b) Incentive mechanism design for crowdsourcing: An all-pay auction approach. *ACM Trans. Intell. Syst. Technol* 7(3):35:1–35:26. <https://doi.org/10.1145/2837029>. ISSN 2157-6904
- Madden S (2004) Intel Lab Data, (accessed June 1, 2017). [Online]. Available: <http://db.csail.mit.edu/labdata/labdata.html>,
- Mchutchon A, Rasmussen CE (2011) Gaussian process training with input noise. In: In Proc. Advances in Neural Information Processing Systems (NIPS), pp 1341–1349
- Minson SE, Brooks BA, Glennie CL, Murray JR, Langbein JO, Owen SE, Heaton TH, Iannucci RA, Darren L. Hauser (2015) Crowdsourced earthquake early Science Advances warning, 1(3)
- Mun M, Reddy S, Shilton K, Yau N, Burke J, Estrin D, Hansen M, Howard E, West R, Péter B (2009) PEIR The personal environmental impact report, as a platform for participatory sensing systems research. In: Proc. ACM Mobile Syst. Applications, and Services (MobiSys), pp 55–68
- Murakami D, Peters GW, Yamagata Y, Matsui T (2016) Participatory sensing data tweets for micro-urban real-time resiliency monitoring and risk management. *IEEE Access* 4:347–372
- Nissim K, Orlandi C, Smorodinsky R (2012) Privacy-aware mechanism design. In: Proc. ACM Conf. Electronic Commerce (EC), pp 774–789
- Overeemn A, Robinson JR, Leijnse H, Steeneveld GJ, Horn BP, Uijlenhoet R (2013) Crowdsourcing urban air temperatures from smartphone battery temperatures. *Geophysical Research Letters* 40(15):4081–4085
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M, Duchesnay E (2011) Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12:2825–2830
- Powell SG, Batt RJ (2008) Modeling for insight: a master class for business analysts. Wiley, Hoboken NJ
- Rasmussen CE, Williams CKI (2005) Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning). MIT Press, Cambridge, MA. ISBN 026218253X
- Restuccia F, Das SK, Payton J (2016) Incentive mechanisms for participatory sensing: Survey and research challenges. *ACM Trans. Sen. Netw* 12(2):13:1–13:40. <https://doi.org/10.1145/2888398>. ISSN 1550-4859
- Rosen JB (1965) Existence and uniqueness of equilibrium points for concave n-person games. *Econometrica* 33(3):520–534
- Shah N, Zhou D (2016) No oops, You won't do it again: Mechanisms for self-correction in crowdsourcing. In: International Conference on Machine Learning, pp 1–10
- Shah N, Zhou D, Peres Y (2015) Approval voting and incentives in crowdsourcing. In: International Conference on Machine Learning, pp 10–19
- Singla A, Krause A (2013) Truthful incentives for privacy tradeoff: Mechanisms for data gathering in community sensing. In: Proc. Int. Conf. Machine Learning (ICML)

- Thiagarajan A, Ravindranath L, LaCurts K, Madden S, Balakrishnan H, Toledo S, Eriksson J (2009) Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones. In: Proc. ACM Conf. Embedded Networked Sensor Syst. (SenSys), pages 85–98, <https://doi.org/10.1145/1644038.1644048>. ISBN 978-1-60558-519-2
- Xiong H, Zhang D, Chen G, Wang L, Gauthier V, Barnes LE (2016) Icrowd: Near-optimal task allocation for piggyback crowdsensing. IEEE Trans. Mobile Comput 15(8):2010–2022. <https://doi.org/10.1109/TMC.2015.2483505>. ISSN 1536-1233
- Yang D, Fang X, Xue G (2013) Truthful incentive mechanisms for k-anonymity location privacy. In: Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM), pages 2994–3002, <https://doi.org/10.1109/INFCOM.2013.6567111>
- Yang D, Xue G, Fang X, Tang J (2012) Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing. In: Proc. ACM Mobile Comput. and Networking (MobiCom), pages 173–184, <https://doi.org/10.1145/2348543.2348567>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.