Privacy-Aware Incentive Mechanism for Mobile Crowd Sensing

Jing Yang Koh*‡, Gareth W. Peters†, Derek Leong‡, Ido Nevat‡ and Wai-Choong Wong*
*Graduate School for Integrative Sciences and Engineering, National University of Singapore, Singapore
†Department of Statistical Science, University College London, London, UK
‡Institute for Infocomm Research (I^{2}R), A*STAR, Singapore

***Abstract*—Mobile crowd sensing is an emerging sensing paradigm where sensing applications buy sensor data from mobile smartphone users (workers) instead of deploying their own sensor networks to estimate some statistics of a spatial event. In many spatial monitoring applications, the crowdsourcer needs to incentivize smartphone users to contribute sensing data so that the collected dataset has good spatial coverage. To further incentivize privacy-concerned workers to contribute, we propose a Stackelberg incentive framework that allows workers to specify their location privacy requirements while also increasing the spatial coverage of the collected dataset. We then derive a unique Stackelberg equilibrium which demonstrates the stability of our approach. Our simulation results show that our approach is significantly better in terms of data utility than the non-location-aware and uniform-reward approaches.**

***Index Terms*—Incentive mechanism design, Stackelberg game, location privacy, mobile crowd sensing.**

I. Introduction

The mobile crowd sensing platform [1] is an emerging sensing paradigm in the age of Internet of Things (IoT) that replaces the fixed sensing infrastructure and removes its deployment and maintenance costs. The sensing platform can exploit the increasing smartphone ownership and incentivize the smartphone users to contribute sensing data that are easily obtained via the available sensing capabilities of their smartphones. This allows the sensing platform to estimate some statistics of a spatial event. However, the sensing platform must first design an appropriate incentive mechanism to encourage user participation, e.g., via monetary rewards.

Many important mobile crowd sensing applications for spatial monitoring such as those used for traffic monitoring [2], earthquake detection [3] or noise monitoring [4] will benefit greatly if the coverage area of its dataset is maximized. Hence, improving the *spatial coverage* of the collected dataset should be one of the main objectives of an incentive mechanism used by spatial monitoring applications. Additionally, current privacy-preserving works such as [5]–[7] 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 the amount of potential data available to the crowdsourcer.

However, the location privacy problem has not been fully addressed as existing schemes that offer location privacy via location or data perturbation are not directly applicable to crowd sensing applications that require specific and true location information. 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.

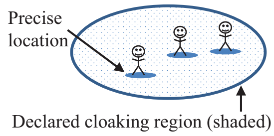


Fig. 1. Privacy model of users: privacy-concerned users can declare a cloaking region that contains their true location instead of providing fine-grained location information to the crowdsourcer.

An appropriate incentive mechanism to model the hierarchical relationship between the crowdsourcer and the smartphone users is the *Stackelberg (leader-follower) game* model used in [8]–[10]. 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 to the crowdsourcer. However, existing Stackelberg incentive models simply select user data independently of their physical location [11] and thus, 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 proposed model allows privacy-sensitive smartphone users to submit *coarse-grained (or quantized) location* information (see Fig. 1) that encompasses their true location (similar to the location cloaking principle in [12], [13]) as the coarse-grained location information could still be useful to the crowdsourcer. We then study the properties of the proposed Stackelberg game analytically and present an efficient algorithmic solution for the Followers' game. Additionally, our proposed model does not require a trusted third party for privacy and can protect users against a crowdsourcer who cannot be trusted to anonymize the users' location information.

A. Related Work

The most common incentive mechanism in the literature is the auction games [5]–[8], [11], [14] where the smartphone workers submit bids (e.g., prices or efforts in [15]) for their data while the crowdsourcer selects a set of workers with the lowest bids and rewards them accordingly. Another fundamentally different game-theoretic approach to model the incentive problem is the Stackelberg game [8]–[10] where the crowdsourcer (leader) first decides on the total reward to pay workers while the workers (followers) then individually decide on the optimal amount of data to contribute.

Privacy-aware incentive mechanisms were considered in [5]–[7] which used data perturbation or dummy locations [16] to protect location privacy. However, the methods may not be applicable to many spatial monitoring applications such as those used for traffic monitoring [2], earthquake detection [3] or even noise monitoring [4] where a dataset with perturbed data or dummy location may trigger a false alarm and make the application unreliable. Thus, the works in [12], [13] used the cloaking region technique (i.e., coarse-grained location information) to provide privacy for workers. In addition, the work in [17] conducted a survey and found that smartphone users were more willing to provide coarse-grained location information than fine-grained information. 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., spatial coverage area) of each worker's data. However, adding location information increases the computational complexity of the problem and thus, [11] proposed an auction-based approximation algorithm to assign 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.

B. Our Contributions

The main contributions of the paper are as follows: 1) We propose a novel user location privacy-aware incentive scheme that allows privacy-sensitive smartphone users to quantize their location information using *cloaking regions* and also improves the spatial coverage of the collected dataset. 2) We formulate the problem as a Stackelberg game, and show the stability (i.e., there exists a unique Stackelberg equilibrium) for our approach. 3) Our numerical analysis shows that our approach achieves significantly better performances in terms of data utility compared to the non-location-aware and uniform-reward approaches.

II. Problem Formulation and Analysis

In this paper, we address the following problem statement: Suppose there is a crowdsourcer (buyer) who aims to buy sensing data from smartphone users (workers), design an incentive mechanism such that the collected dataset has (i) good *spatial coverage*, and (ii) is *location privacy-preserving* for the workers. We first present our system model and the proposed Stackelberg framework before analyzing its properties.

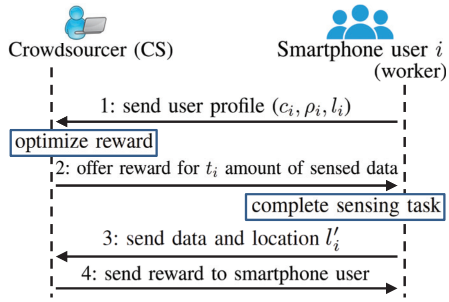


Fig. 2. Interaction model between the crowdsourcer and smartphone users (workers). The user profile consists of the user's sensing cost incurred per bit of data c_i, privacy coefficient ρ_i, and cloaking region l_i.

A. System Model and Assumptions

The crowd sensing system consists of a set of $\mathcal{I} = \{1, \ldots, N\}$ workers and a single crowdsourcer who partitions the entire spatial area of interest into a set of L cloaking 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 $c_i \in (0, c^{\max}]$, privacy coefficient $\rho_i \in [0, \rho^{\max}]$, location l'_i (which may be of different granularity for each worker) and its corresponding cloaking region l_i defined by the system where $l'_i \in l_i$ and $l_i \in \mathcal{L}$. The interaction model between the crowdsourcer and 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 his utility. The crowdsourcer then offers the selected workers a reward in exchange for some amount of sensing data. Next, the selected workers will collect their data and transmit them along with their locations to the crowdsourcer and receive their rewards. Note that there are two cloaking regions l_i and l'_i in our model, l_i is the initial (coarse-grained) cloaking region defined by the system and l'_i is the worker's own declared cloaking region, which is submitted only when the worker is selected.

We assume that the granularity of the worker's submitted location l'_i is inversely proportional to the privacy coefficient ρ_i since a privacy-sensitive worker is likely to provide only coarse-grained location information. Hence, the parameter ρ_i allows the crowdsourcer to differentiate workers and preserve the privacy of unselected workers since they only reveal their locations when they are selected. Note that the workers can anonymize their precise location information via cloaking regions and do not rely on the crowdsourcer for anoymization. In practice, this can be implemented in the crowdsourcing software in the workers' devices to allow selection of different cloaking regions with (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) as the *followers*. The crowdsourcer acts first and commits a reward strategy while the workers subsequently choose their best response after observing the crowdsourcer's strategy. The strategy of the crowdsourcer is the reward for each cloaking region $\boldsymbol{R} = (R_1, \ldots, R_L)$ and the strategy of worker i is the amount of data to contribute (in terms of sensing time) $t_i \geq 0$.

The crowdsourcer only optimizes the reward R_l allocated to each cloaking region $l \in \mathcal{L}$ and subsequently offers each participating worker i a fraction of R_l depending on the proportion of its contribution and its location:

$$\text{Worker } i\text{'s offered reward} = \frac{t_i}{\sum\limits_{j \in \mathcal{Q}: l_j = l_i} t_j} R_{l_i}, \tag{1}$$

where $\mathcal{Q}$ is the set of participating workers i with $t_i > 0$. We assume that $|j \in \mathcal{Q} : l_j = l_i| > 1$.

B. Utility Functions for Crowdsourcer and Workers

Crowdsourcer: We define the utility function of the crowdsourcer to be $U_{CS}(\boldsymbol{R}; \boldsymbol{t}) = \sum_{i \in \mathcal{I}} f_d(t_i, l_i)$, where $f_d(t_i, l_i)$ is a function of the quantity and quality (e.g., spatial coverage and location granularity) of the data from each worker i. For simplicity, we let $f_d(t_i, l_i) = \lambda_i \log(1 + t_i)$, where $\lambda_i \geq 0$ is a system parameter and the log function is used to model the diminishing returns [8] (in addition to its concavity, for mathematical convenience) on each worker's data. Thus, the crowdsourcer's utility function is given by

$$U_{CS}(\boldsymbol{R}; \boldsymbol{t}) = \sum_{i \in \mathcal{I}} \lambda_i \log(1 + t_i). \tag{2}$$

To increase the coverage area of the worker's dataset, a higher λ_i can be assigned to workers located at less populated regions. In addition, a higher λ_i can be assigned to workers who provide finer location information[1], which is assumed to provide higher utility for the crowdsourcer. By introducing the λ_i parameters, the crowdsourcer is able to differentiate between the quality (e.g., spatial coverage) of each worker's data.

Workers: The utility function of worker i is defined to be the amount of reward it receives from the crowdsourcer as defined in (1) minus the cost incurred for obtaining the data: $u_i(t_i; \boldsymbol{t_{-i}}, R_l) = \frac{t_i}{\sum_{j \in \mathcal{Q}: l_j = l_i} t_j} R_l - f_c(c_i, \rho_i) t_i$, where $f_c(c_i, \rho_i)$ is a cost function, $t_i \geq 0$ is the amount of data (in terms of sensing time) sold to the crowdsourcer, $\boldsymbol{t_{-i}}$ is a vector of the amount of data sold by all workers except worker i. For simplicity, we assume that the workers do not incur any cost in exchanging their user profiles (stage 1 of Fig. 2). We let $f_c(c_i, \rho_i) = c_i(1 + \rho_i)$, where ρ_i has a multiplicative effect on the worker's overall cost. Therefore, worker i's utility function is given by

$$u_i(t_i; \boldsymbol{t_{-i}}, R_l) = \frac{t_i}{\sum\limits_{j \in \mathcal{Q}: l_j = l_i} t_j} R_l - c_i(1 + \rho_i) t_i. \tag{3}$$

C. Stackelberg Game Formulation

Given that the crowdsourcer wants to increase the coverage area of his dataset while satisfying a budget constraint R^{budget} and a minimal reward allocation $R_l^{\min} > 0$ for each region l (this allows the crowdsourcer to specify more important regions), it solves the following optimization problem:

Problem 1 $(\boldsymbol{\lambda}, \boldsymbol{R}^{\mathbf{min}}, R^{\text{budget}})$

$$\begin{aligned} \underset{\boldsymbol{R}}{\text{maximize}} \quad & \sum_{i \in \mathcal{I}} \lambda_i \log(1 + t_i) \\ \text{subject to} \quad & R_l \geq R_l^{\min} \qquad , \forall l \in \mathcal{L}, \\ & \sum_{l \in \mathcal{L}} R_l \leq R^{\text{budget}}, \end{aligned} \tag{4}$$

where t_i is the optimal solution to Problem 2.

Each worker i solves the following optimization problem:

Problem 2 $(i, \boldsymbol{t_{-i}}, R_l, c_i, \rho_i)$

$$\begin{aligned} \underset{t_i}{\text{maximize}} \quad & \frac{t_i}{\sum\limits_{j \in \mathcal{Q}: l_j = l_i} t_j} R_l - c_i(1 + \rho_i) t_i, \\ \text{subject to} \quad & t_i \geq 0. \end{aligned} \tag{5}$$

Problems 1 and 2 form a Stackelberg game and our goal is to find the Stackelberg equilibrium point 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 its strategy.

Definition 1 (Stackelberg Equilibrium)**.** *Let $\boldsymbol{R}^*$ be the optimal solution for the crowdsourcer, obtained by solving Problem 1 and $\boldsymbol{t}^*$ be the optimal solution for the workers, obtained by solving Problem 2. The strategy profile $(\boldsymbol{R}^*, \boldsymbol{t}^*)$ is a Stackelberg equilibrium for the proposed Stackelberg game if the following conditions are satisfied for any $(\boldsymbol{R}, \boldsymbol{t})$ where $\boldsymbol{R} \succeq 0, \boldsymbol{t} \succeq 0$:*

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

We apply the backward induction method to analyze the proposed Stackelberg game. First, we start with the Followers game and study the predicted best response t_i^* for each worker i as a function of the reward R_l offered by the crowdsourcer and the strategies of the other workers $\boldsymbol{t_{-i}}$. Subsequently, we analyze the best response of the crowdsourcer in Problem 1.

D. Nash Equilibrium of Followers Game

We first consider the Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$ where $\mathcal{I}$ is the player set of N workers and u_i is the utility function of worker i. We show in Theorem 1 that there exists a unique Nash equilibrium in the Followers game.

Lemma 1. *A Nash equilibrium exists in the Followers game $(\mathcal{I}, \{t_i\}_{i \in \mathcal{I}}, \{u_i\}_{i \in \mathcal{I}})$. See Appendix-A for proof.*

To study the best response strategy for worker i given the strategies of the other players $\boldsymbol{t_{-i}}$, we set $\frac{\partial u_i}{\partial t_i} = 0$:

$$\sum_{j: l_j = l_i, j \neq i} t_j R_l = c_i(1 + \rho_i) \left(\sum_{j: l_j = l_i, j \neq i} t_j + t_i \right)^2. \tag{6}$$

We now seek an expression for the optimal t_i^* value that is independent of the other t_j^* values. We say that a worker i is a participating worker if $t_i^* > 0$.

[1] For example, in our simulation Section III, we let $\lambda_i = 1 + 15/\log(1 + N_l)$ if $\rho_i > \rho^{\max}/2$ and $\lambda_i = 0.75 + 15/\log(1 + N_l)$, otherwise.

Theorem 1. *The Followers game given by the triplet $(\mathcal{I}, \{t_i\}_{i\in\mathcal{I}}, \{u_i\}_{i\in\mathcal{I}})$ has a unique Nash equilibrium given by the following closed-form expression:*

$$t_i^* = \begin{cases} \frac{(|\mathcal{Q}_{l_i}|-1)R_l}{\sum\limits_{j\in\mathcal{Q}_{l_i}} c_j(1+\rho_j)}\Big(1-\frac{(|\mathcal{Q}_{l_i}|-1)c_i(1+\rho_i)}{\sum\limits_{j\in\mathcal{Q}_{l_i}} c_j(1+\rho_j)}\Big), \\ \qquad \text{if } i\in\mathcal{Q}_{l_i}, \\ 0, \qquad \text{otherwise,} \end{cases} \tag{7}$$

where $\mathcal{Q}_{l_i}$ is the set of participating workers in region l_i. See Appendix-B for proof.

Note that $t_i^* > 0$ for all participating workers i in (7). Hence, for all participating workers i, the following inequality should be satisfied: $1-\frac{(|\mathcal{Q}_{l_i}|-1)c_i(1+\rho_i)}{\sum_{j\in\mathcal{Q}_{l_i}} c_j(1+\rho_j)} > 0$. This leads to the following condition, which will be used in our Algorithm 1.

$$c_i(1+\rho_i) < \frac{\sum\limits_{j\in\mathcal{Q}_{l_i}} c_j(1+\rho_j)}{|\mathcal{Q}_{l_i}|-1} \quad ,\forall i\in\mathcal{Q}_{l_i}. \tag{8}$$

The optimal t_i^* for the workers is given by the Nash equilibrium solution (7) of the Followers game. However, (7). requires knowledge of the set of participating workers $\mathcal{Q}_{l_i}$, c_i, and ρ_i. Hence, we propose Algorithm 1 (adapted from [8]) which makes use of (8) to greedily compute $\mathcal{Q}_{l_i}$ and solve for t_i^*. It can be easily shown that Algorithm 1 is correct as all participating workers are in equilibrium, i.e., $\frac{\partial u_i}{\partial t_i} = 0$, $\forall i\in\mathcal{Q}$.

E. Stackelberg Equilibrium

Using the analytical result (7) for the Followers game, the crowdsourcer can optimize his reward strategy $\boldsymbol{R}$ efficiently by substituting (7) into his utility function in (2). By Theorem 2, there exists a unique Stackelberg equilibrium which results in a stable equilibrium strategy profile. This allows the crowdsourcer to predict the behaviors of the workers and efficiently compute $\boldsymbol{R}^*$. From (7), it can be observed that the equilibrium solution of a participating worker i is inversely proportional to both c_i and ρ_i. Both Theorems 1 and 2 extend [8, Theorem 2] to the case where the workers' privacy coefficient ρ and location l are also considered.

Theorem 2. *The proposed Stackelberg game has a unique Stackelberg equilibrium.*

Proof. Recall from Theorem 1 that the Followers Game has a unique Nash equilibrium. It can be easily shown that the best response strategy set of the crowdsourcer is convex and compact since R_l is assumed to be bounded, and U_{CS} is continuous in $\boldsymbol{R}$. Hence, we need to show the concavity of U_{CS} to conclude that there exists a unique Stackelberg equilibrium. The second-order partial derivatives of U_{CS} with respect to R_l and R_lR_k are as follow:

$$\frac{\partial^2 U_{CS}}{\partial R_l^2} = -\sum_{i:l_i=l}\lambda_i\Big(\frac{\tau_i^2}{(\tau_i R_l+1)^2}\Big) < 0,\ \frac{\partial^2 U_{CS}}{\partial R_l \partial R_k} = 0, \tag{9}$$

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

Algorithm 1: Compute Nash equilibrium for Followers game $(\mathcal{I}, \{t_i\}_{i\in\mathcal{I}}, \{u_i\}_{i\in\mathcal{I}})$ played by the workers.

function SolveFollowersGame($\boldsymbol{c}, \boldsymbol{\rho}, \boldsymbol{l}, \boldsymbol{R}$)
Input : sensing costs $\boldsymbol{c}_{1,\dots,N}$, privacy coefficients $\boldsymbol{\rho}_{1,\dots,N}$, workers' cloaking regions $\boldsymbol{l}_{1,\dots,N}$, rewards $\boldsymbol{R}_{1,\dots,L}$.
Output: data sold to crowdsourcer $\boldsymbol{t}^*_{1,\dots,N}$.
foreach *region* $l\in\mathcal{L}$ **do**
  Sort the group of workers in l according to their privacy-weighted unit cost in ascending order where $c_i(1+\rho_i) \le c_{i+1}(1+\rho_{i+1})$.
  Let $\mathcal{Q}_l = \{1,2\}$ be the set of participating workers with $t_i > 0$.
  Set $\mathcal{Q}_l \leftarrow \mathcal{Q}_l \cup \{i\}$ for each worker i in region l if $c_i(1+\rho_i) < \frac{1}{|\mathcal{Q}_l|-1}\sum_{j\in\mathcal{Q}_l} c_j(1+\rho_j)$ (note: the looping can stop at the ith step when the condition is not met).
  Set t_j^* according to (7) for all j satisfying $l_j = l$.
end

Since the Hessian matrix of U_{CS} is a diagonal matrix, its eigenvalues, i.e., $\frac{\partial^2 U_{CS}}{\partial R_l^2}$ are easily shown to be strictly negative. This implies that the Hessian matrix is negative definite and thus, U_{CS} is strictly concave in $\boldsymbol{R}$. □

By Theorem 2, the optimal reward $\boldsymbol{R}^*$ has a unique maximizer and thus, can be efficiently computed using the well-known interior point methods available in most commercial solvers.

III. Simulations

We use numerical simulations to study the effects of certain parameters on the Stackelberg equilibrium and also compare our proposed scheme against two baseline schemes: (B1) a non-location-aware Stackelberg game, and (B2) a simple uniform-reward scheme. In B1, the crowdsourcer treats all workers as coming from one single location when solving the Stackelberg game, and in B2, the total rewards are distributed equally among all regions containing at least two workers. We used $N = 1,000$ workers and ran the simulation for at least 100 iterations to obtain the average results. The optimal solution for our proposed scheme and scheme (i) were obtained using the CVXPY [18] solver. We simulate five scenarios where the worker's sensing cost c_i, privacy coefficient ρ_i, and cloaking region l_i all come from different parametric distributions. We let $\lambda_i = 1 + 15/\log(1+N_l)$ if $\rho_i > \rho^{\max}/2$ and $\lambda_i = 0.75 + 15/\log(1+N_l)$, otherwise, where N_l is the number of workers in region l, $c^{\max} = 5$, $\rho^{\max} = 5$, $R_l^{\min} = 0.1, \forall l\in\mathcal{L}$, and $R^{\text{budget}} = 300$. For simplicity, we only distinguish between two levels of ρ_i as they can represent the coarse and fine granularity of the workers' location information.

A. Simulation Scenarios

Scenario (S1 - uniform): We let the worker's sensing cost c_i, privacy coefficient ρ_i, and cloaking regions l_i be chosen uniformly at random from $[1, c^{\max}]$, $[0, \rho^{\max}]$, and $[1, L]$ respectively. **Scenario (S2 - privacy-sensitive):** We use the same c_i and l_i distribution as S1 but let $\rho_i = \rho^{\max}/N_l$.

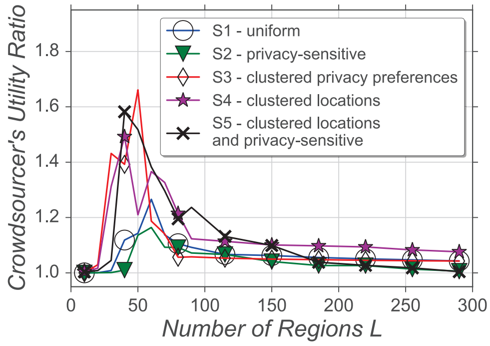


Fig. 3. Ratio of CS utility (2) from the proposed scheme over the utility obtained from non-location-aware scheme B1. Note that L starts from 10.

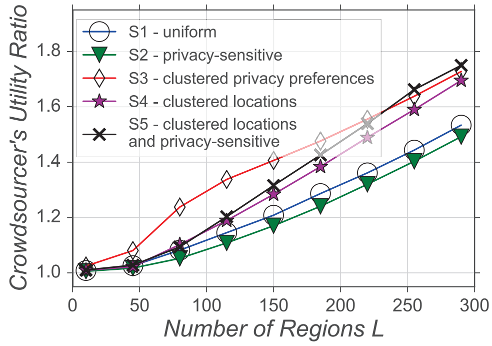


Fig. 4. Ratio of CS utility (2) from the proposed scheme over the utility obtained from uniform-reward scheme B2. Note that L starts from 10.

The intuition behind is that workers at "less populated" regions may be more privacy sensitive as they may be easily identified. **Scenario (S3 - clustered privacy preferences):** We use the same c_i and l_i distribution as S1 but let the ρ_i values in each region be randomly chosen from a Gaussian mixture distribution with standard deviation of 0.1 with 20% of the points having a mean of 1 and the rest with a mean of 3. The mixture model is motivated by an earlier study [17]. **Scenario (S4 - clustered locations):** We use the same c_i and ρ_i distribution as S1 but consider the case where the worker's locations are clustered and randomly chosen from a non-homogeneous binomial point process where 50% of them are located in one-fifth of the regions while the rest are uniformly distributed in the other four-fifths. **Scenario (S5 - clustered locations and privacy-sensitive):** Finally, we combined S2 and S4 and let c_i be uniformly distributed, $\rho_i = \rho^{\max}/N_l$, and l_i is randomly chosen as in S4.

B. Comparison Against the (B1) Non-Location-Aware and (B2) Uniform-Reward Schemes

We plot the ratio of the crowdsourcer (CS) utility (2) from the proposed scheme over the utility of the non-location-aware scheme (i.e., $U_{CS}^{\text{proposed}}/U_{CS}^{\text{baseline}}$) in Fig. 3. The ratio is always above 1 which indicates that the proposed scheme is always better or as good as B1. Our proposed scheme outperforms B1 especially when $L = 50$. When $L < 50$, there are a large number of workers in each region and hence, the two schemes were able to find similarly cheap workers. But as $L > 70$, the number of workers in each region deceases and there exists regions with < 2 workers. Hence, the improvements start to decrease. Interestingly, the improvement in S5 is more significant than S4 initially when $L < 200$ but becomes less than S4 as L becomes larger. This is because the workers' costs are much cheaper in the clustered areas in S5 when L is small. As such, more rewards were allocated to the clustered areas. Next, we plot the ratio of the CS utility (2) from the proposed scheme over the utility of the uniform-reward scheme in Fig. 4. The performance of our scheme is significantly better than B2 when $L > 150$, especially in scenarios S3, S4 and S5 where the distributions are clustered. In general, the improvement in the CS utility increases as L increases. This is because as there are lesser workers in each region, the probability of finding a cheaper worker also decreases compared to when there are many workers in the region. Thus, B2 does not work well as it does not allocate more rewards to regions with cheaper workers.

Finally, we study how the proposed scheme allocates the rewards under the five different scenarios when $L = 150$ and plot the results in Fig. 5 for a single instantiation. We observed the following key points (which are typical across multiple realizations). Due to the coverage increasing nature of the chosen λ_i parameters, the number of participating workers (in orange) tend to be uniformly distributed, although more reward is allocated to regions with cheaper workers in most cases, e.g., in S3, S4, and S5 where the workers in populated regions are less privacy-sensitive. The most non-uniform reward allocation occurs in S3 where the privacy preferences of the workers are clustered and only a small number of regions have relatively cheaper workers.

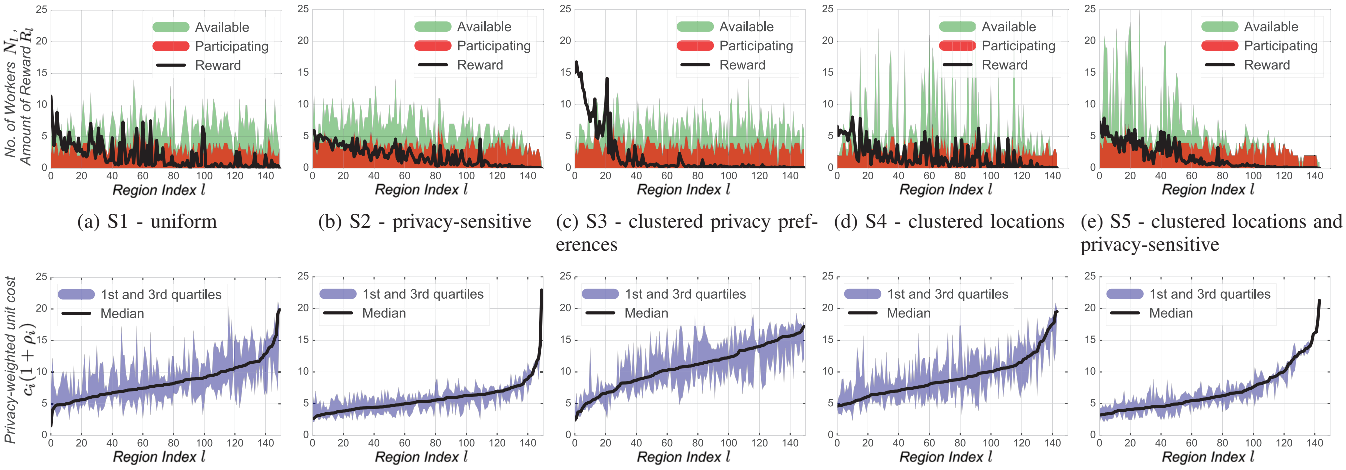
IV. Conclusion

This paper proposes a Stackelberg framework for mobile crowd sensing applications to incentivize privacy-sensitive smartphone users while increasing the coverage of the dataset. Our approach offers stronger incentives to privacy-sensitive participants by allowing them to use cloaking regions to hide their precise location. We considered different crowdsourcing environments and analyzed their influence on the Nash equilibrium point using simulations. Our simulation results show that our model is significantly better than the non-location-aware and uniform-reward schemes in terms of data utility. Our future work will be to consider uncertainty and untruthfulness in the users' privacy coefficients and sensing costs.

Appendix

A. Proof of Lemma 1

A Nash equilibrium (NE) exists in the Followers game if: for all $i \in \mathcal{I}$, (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 concave in t_i [19]. 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 [see (3)], and concave in t_i since



(a) S1 - uniform (b) S2 - privacy-sensitive (c) S3 - clustered privacy preferences (d) S4 - clustered locations (e) S5 - clustered locations and privacy-sensitive

Fig. 5. Reward allocation per region and worker participation distribution (top row) and worker privacy-weighted unit cost $c_i(1+\rho_i)$ distribution (bottom row) under different scenarios. The regions are sorted accordingly to the median privacy-weighted cost in ascending order for each scenario.

the second-order partial derivative of u_i with respect to t_i is strictly negative: $\frac{\partial^2 u_i}{\partial t_i^2} = \frac{-2 \sum_{j:l_j=l_i, j\neq i} t_j}{\left(\sum_{j:l_j=l_i} t_j\right)^3} R_l < 0.$ □

B. Proof of Theorem 1

Since the utility function $u_i(t_i, R_l)$ of each worker i is strictly concave in t_i as shown in Lemma 1, there exists a unique strategy that maximizes the utility of each worker given the strategies of the other players. To compute the unique NE point t_i^* for each worker i, we need to solve (6) for all participating workers. First, we manipulate (6) to obtain t_i^*:

$$t_i^* = \Big(\sum_{j:l_j=l_i} t_j\Big)\Big(1 - c_i(1+\rho_i)\Big(\sum_{j:l_j=l_i} t_j\Big)R_l^{-1}\Big). \tag{10}$$

Next, we sum up the t_i^* in (10) for all participating workers (with $t_i > 0$) in region l_i (denoted by $\mathcal{Q}_{l_i}$) to obtain:

$$\sum_{j\in\mathcal{Q}_{l_i}} t_j^* = \frac{(|\mathcal{Q}_{l_i}|-1)R_l}{\sum_{j\in\mathcal{Q}_{l_i}} c_j(1+\rho_j)}. \tag{11}$$

Finally, we substitute (11) into (10) to obtain the unique NE point for each worker i given by (7). □

References

[1] B. Guo, Z. Wang, Z. Yu, Y. Wang, N. Y. Yen, R. Huang, and X. Zhou, "Mobile crowd sensing and computing: The review of an emerging human-powered sensing paradigm," *ACM Comput. Surv.*, vol. 48, pp. 7–31, Aug. 2015.

[2] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones," in *Proc. ACM Conf. Embedded Networked Sensor Syst. (SenSys)*, pp. 85–98, 2009.

[3] S. E. Minson, B. A. Brooks, C. L. Glennie, J. R. Murray, J. O. Langbein, S. E. Owen, T. H. Heaton, R. A. Iannucci, and D. L. Hauser, "Crowdsourced earthquake early warning," *Science Advances*, vol. 1, Apr. 2015.

[4] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in *Proc. ACM/IEEE Conf. Inf. Process. in Sensor Netw. (IPSN)*, pp. 105–116, 2010.

[5] K. Nissim, C. Orlandi, and R. Smorodinsky, "Privacy-aware mechanism design," in *Proc. ACM Conf. Electronic Commerce (EC)*, pp. 774–789, 2012.

[6] D. Yang, X. Fang, and G. Xue, "Truthful incentive mechanisms for k-anonymity location privacy," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, pp. 2994–3002, 2013.

[7] A. Singla and A. Krause, "Truthful incentives for privacy tradeoff: Mechanisms for data gathering in community sensing," in *ICML Workshop: Machine Learning Meets Crowdsourcing*, Jun. 2013.

[8] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. ACM Mobile Comput. and Networking (MobiCom)*, pp. 173–184, 2012.

[9] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, pp. 1701–1709, 2012.

[10] S. Luo, Y. Sun, Y. Ji, and D. Zhao, "Stackelberg game based incentive mechanisms for multiple collaborative tasks in mobile crowdsourcing," *Mobile Networks and Applications*, vol. 21, no. 3, pp. 506–522, 2016.

[11] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, pp. 1231–1239, Apr. 2014.

[12] R. Cheng, Y. Zhang, E. Bertino, and S. Prabhakar, "Preserving user location privacy in mobile data management infrastructures," in *Int. Workshop on Privacy Enhancing Technologies (PET)*, pp. 393–412, Jun. 2006.

[13] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. ACM Conference on Computer (CCS)*, pp. 901–914, 2013.

[14] F. Restuccia, S. K. Das, and J. Payton, "Incentive mechanisms for participatory sensing: Survey and research challenges," *ACM Trans. Sen. Netw.*, vol. 12, pp. 13:1–13:40, Apr. 2016.

[15] T. Luo, S. K. Das, H. P. Tan, and L. Xia, "Incentive mechanism design for crowdsourcing: An all-pay auction approach," *ACM Trans. Intell. Syst. Technol.*, vol. 7, pp. 35:1–35:26, Feb. 2016.

[16] H. Kido, Y. Yanagisawa, and T. Satoh, "An anonymous communication technique using dummies for location-based services," in *Proc. Int. Conf. Pervasive Services.*, pp. 88–97, Jul. 2005.

[17] J. Lin, B. Liu, N. Sadeh, and J. I. Hong, "Modeling users' mobile app privacy preferences: Restoring usability in a sea of permission settings," in *Symp. Usable Privacy and Security (SOUPS)*, Jul. 2014.

[18] S. Diamond and S. Boyd, "CVXPY: A Python-embedded modeling language for convex optimization," *J. Machine Learning Research*, vol. 17, no. 83, pp. 1–5, 2016.

[19] J. B. Rosen, "Existence and uniqueness of equilibrium points for concave n-person games," *Econometrica*, vol. 33, no. 3, pp. 520–534, 1965.