Incentive Design for Air Pollution Monitoring based on Compressive Crowdsensing

Tong Liu†, Yanmin Zhu†, Yuanyuan Yang§, Fan Ye§

†Shanghai Jiao Tong University; §Stony Brook University

†{liutong25691640, yzhu}@sjtu.edu.cn; §{yuanyuan.yang, fan.ye}@stonybrook.edu

***Abstract*—As air pollution is becoming a serious problem in developing nations, governments try to track and solve this problem by monitoring air pollution. With the proliferation of smartphones, mobile crowdsensing becomes a promising paradigm for monitoring fine-grained air pollution in urban areas. As existing studies have shown that pollutant concentrations have inherent spatiotemporal correlations, compressive sensing is an effective technology to reduce the amount of data collected through crowdsensing. In a practical crowdsensing application, incentives are expected by smartphone users for contributing sensing data. However, how to design incentives to collect high-quality sensing data with low costs is difficult in compressive crowdsensing. In this work, we propose an iterative scheme for the process of crowdsensing-based air pollution monitoring, where incentives are updated online according to the distribution of collected sensing data. Comprehensive simulations have been conducted to demonstrate the efficacy of our proposed scheme.**

I. Introduction

With the modernization of people's life, air pollution emerges as an acute problem in urban areas, especially developing nations (e.g., China and India). Long-term exposure in air pollutants such as NO_2, $PM_{2.5}$ and CO, leads to a high risk of several health problems, such as respiratory infections, heart disease and lung cancer. To track and solve this problem, governments try to monitor air pollution, by deploying fixed measuring stations. However, building such stations is severely limited by the availability of land and high costs of maintenance (about \$30,000 per year), which results in impossibility to obtain fine-grained pollution measurements in a large urban area. For example, only 22 measuring stations are located in Beijing, an area of 16400km^2. Although several works [1] have focused on inferring fine-grained air qualities by exploiting the correlations with other datasets such as traffic flows and points of interest, direct measurements provide fidelity and accuracy unsurpassable by other means.

Fortunately, crowdsensing provides an unprecedented opportunity for collecting sensing data in a large scale (e.g., community or city), which takes advantage of widely-distributed mobile sensing devices such as smartphones. Numerous environment-centric applications have been developed based on crowdsensing, such as traffic monitoring and noise mapping. In these applications, smartphone users report their location-based measurements to a central platform via wireless networks. After aggregating plenty of widely-distributed measurements, the platform can obtain the overview of an environmental phenomenon. Similarly, air pollution monitoring can be conducted based on crowdsensing. Although smartphones are not equipped with environmental sensors at present, fortunately, sensor-integrated portable external hardware [2] has been developed. Moreover, smartphones incorporating small low-cost environmental sensors are coming soon [3].

Recently, several efforts [4][5][6] have been put into developing crowdsensing systems for monitoring air pollution. However, most of them focus on the implementation of portable sensing devices or smartphone applications. For example, a crowdsensing system, named P-Sense, is designed in [5], where external sensing devices are developed to measure the concentrations of several pollutants. Different from these works, we focus on incentive design for a crowdsensing-based air pollution monitoring system, assuming smartphones can measure air pollutant concentrations.

An efficient and appropriate incentive mechanism is a key component in practical crowdsensing systems. On one hand, certain monetary rewards are expected by smartphone users to participate in sensing, because numerous resources are consumed such as energy, bandwidth and human efforts. On the other hand, the platform of a crowdsensing system wants to minimize its total payment under the condition of guaranteeing the quality and quantity of collected data. Although a number of incentive mechanisms [7][8][9][10] have been proposed, they always focus on ensuring the truthfulness of participants. They do not consider the various values of sensing data from an overall perspective and handle each measurement through balancing its value and cost.

To monitor fine-grained air pollution in an urban area, a large amount of measurements in different locations are needed, which is still costly for the platform. Fortunately, inherent spatiotemporal correlations among air pollution data have been observed in real datasets [1], because air pollutants released by pollution sources disperse in space following certain laws (e.g., the Gaussian model [11]). By exploiting the correlations, *compressive sensing* (CS) technology [12] can be employed to significantly reduce the number of needed measurements and accurately recover the whole pollution map. A few existing works [13][14][15] have combined compressive sensing and crowdsensing into *compressive crowdsensing*. However, these works simply assume smartphone users are cooperative, who will do sensing tasks allocated to them without incentives. In contrast, we try to design an incentive mechanism for a compressive crowdsensing system, which provides location-dependent incentives to participants via balancing the value and cost of each measurement.

978-1-5090-1328-9/16/\$31.00 ©2016 IEEE

In this paper, we consider the problem of incentive design for a practical compressive crowdsensing system used to monitor air pollution in urban areas. In this system, the platform aims at minimizing the total payment paid for participants and detecting the whole pollution map accurately based on collected measurements. This problem is highly difficult due to several challenges. *First*, the relationship between an arbitrary incentive and the participation of smartphone users is not clear. It is impractical to collect cost information of all users and then choose the cheapest ones, as it consumes a lot of time and energy. *Second*, the quality of measurements are not guaranteed due to various measuring errors of different smartphones. *Third*, the value of a measurement in a specific location is unknown in terms of recovering the whole pollution map via compressive sensing.

To address these challenges, we first build a probabilistic model to analyze the participation of a crowd of rational smartphone users given certain incentives. Then, we propose an iterative scheme for the crowdsensing process. In each iteration, smartphone users participate in sensing given location-dependent incentives. The platform first estimates air qualities in the locations with sufficient collected measurements and then detects the whole pollution map based on the estimations via compressive sensing. We also provide an algorithm to update location-dependent incentives, which are adaptive to the current collection and detection. Finally, we perform comprehensive simulations to evaluate our scheme in terms of the total payment and the accuracy of detection.

The remainder of this paper proceeds as follows. The models and preliminaries are presented in Section II. Section III describes the details of our iterative scheme for crowdsensing-based air pollution monitoring and the algorithm proposed for incentive design. Section IV shows the performance of our simulations. Finally, we discuss related work and conclude the paper in Section V and Section VI, respectively.

II. Models and Preliminaries

A. System model

In an urban area, air pollutants are always released by several natural or anthropogenic pollution sources, such as power plants, wild fires and traffic flows. For simplicity, we only consider stationary pollution sources in this paper, leaving mobile sources in our future work. And we assume pollution sources release pollutants continuously at certain emission rates. We suppose there are k pollution sources in total, and their emission rates are denoted by $\mathbf{Q} = \{Q_1, \cdots, Q_k\}$. Note that the number of pollution sources k is unknown in prior.

For the convenience of studying fine-grained air pollution, we virtually divide the whole urban area into n small grids of the same size, e.g., $200\text{m} * 200\text{m}$. The air quality in each grid can be seen as uniform while different grids may have different values. Supposing each pollution source is located in a grid, its location can be denoted by a grid. Note that compared with grids, pollution sources are sparse, which means $k \ll n$. We use a vector $\mathbf{g} = \{g_1, \cdots, g_n\}^T$ to indicate whether a grid contains a pollution source. The value of $\mathbf{g}$ is defined as

$$g_i = \begin{cases} Q_j, \text{ if pollution source } j \text{ is in grid } i \\ 0, \text{ if there is no source in grid } i \end{cases}, \forall 1 \leq i \leq n.$$

Air pollutants released by these sources disperse in space, vitally influenced by winds, which leads to different pollutant concentrations in different locations. We consider the pollutant concentration in a grid as the accumulation of pollutants from all sources dispersed to the grid. In this paper, we apply a classical air pollution dispersion model, which is described in Section II-B. Given the dispersion model, we can derive a transition matrix $\mathbf{\Omega} = \{\Omega_{ij}\} \in \mathbb{R}^{n*n}$, which satisfies

$$\begin{bmatrix} C_1 \\ \vdots \\ C_n \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \cdots & \Omega_{1n} \\ \vdots & \ddots & \vdots \\ \Omega_{n1} & \cdots & \Omega_{nn} \end{bmatrix} * \begin{bmatrix} g_1 \\ \vdots \\ g_n \end{bmatrix},$$

where Ω_{ij} denotes the pollutants in grid i caused by the source in grid j. Here, $\mathbf{C} = \{C_1, \cdots, C_n\}^T$ denotes pollution concentrations in grids, which can be measured by smartphones.

In our scheme, the platform provides the same reward for each measurement in one grid, while rewards for different grids can differ. We denote the reward for each measurement in grid i as r_i, and reward vector $\mathbf{R} = \{r_1, \cdots, r_n\}$. Given a certain reward r_i, some smartphone users in grid i will actively participate in sensing to earn money. To understand the relation between reward r_i and users' participation, we build a probabilistic participation model for a crowd of rational smartphone users (refer to Section II-C for details). We assume the set of collected measurements in grid i is $\mathbf{M}_i = \{m_1^{(i)}, \cdots, m_{\gamma_i}^{(i)}\}$, where γ_i represents the number of measurements.

B. Air pollution dispersion model

Air pollution dispersion modeling uses mathematical equations to simulate the movement of pollutants in atmosphere and predict future concentrations. Several different types of air pollution dispersion models [16] have been proposed. In this paper, we choose a most widely used one, Gaussian model, for our study. Basically, it assumes pollutant concentrations decay according to the Gaussian distribution. Here, we consider pollutants disperse in two-dimensional space for simplification. Accordingly, the *Complete Equation for Gaussian Dispersion Modeling* can be expressed as

$$C = \alpha \cdot \frac{Q}{\nu} \cdot \left(\frac{1}{\beta\sqrt{2\pi}} e^{\frac{-d^2}{2\beta^2}} \right),$$

where Q is the source pollutant emission rate, ν represents the wind velocity, d is the crosswind distance from the source and C denotes the pollutant concentration. α and β are two constant parameters, which measure the atmospheric turbulence. The pollutant concentrations in downwind grids are influenced by the wind velocity and decay along with the crosswind distance.

Assuming the wind velocity and direction can be known from other datasets like weather forecasts, we can calculate the value of transition matrix $\mathbf{\Omega}$ according to the dispersion model as follows,

$$\Omega_{ij} = \frac{\alpha}{\sqrt{2\pi}\beta\nu} \cdot e^{\frac{-d_{ij}^2}{2\beta^2}}, \text{if grid } i \text{ is downwind from } j, \quad (1)$$

where d_{ij} denotes the crosswind distance between grids i and j. In this paper, we assume the emission rates of pollution sources, wind velocity and direction do not change significantly in a certain period. Therefore, the air pollution concentrations remain stable as well.

C. User participation model

In this subsection, we *first* build a model to characterize the participation behavior of one user given incentives. As some resources (e.g., energy and bandwidth) are consumed for measuring pollution concentrations, a certain amount of costs are incurred. Intuitively, a rational user will not participate in sensing if the reward he/she earns is less than the cost. Thus, we apply a straightforward model, in which the participation decision of user s is denoted by a random variable X_s and there exists

$$X_s = \begin{cases} 0, & \text{if reward } r \text{ is less than cost } c_s \\ 1, & \text{otherwise} \end{cases}.$$

The cost of a specific user depends on many factors, such as the hardware of smartphones, the remaining energy of batteries and the quality of wireless networks. These factors lead to various costs on different users, which is private information of each user.

As it is unnecessary to collect measurements from all users, the platform tends to choose users with the lowest costs for the sake of saving money. However, it is impossible to collect the cost information from all users and then pick up the cheapest ones. Firstly, the process of collecting information from all users is costly in both latency and energy. Secondly, the smartphone users, who submit their cost information without being chosen, will lose interests in participating in the future. To avoid collecting the cost information, we set a uniform reward r_i for each measurement collected in grid i. Thus, only users with $c_s \le r_i$ will participate in sensing. The extra money $(r_i - c_s)$ paid can be seen as the cost of avoiding collecting private information from all users. The probability distribution of X_s can be represented as

$$f(X_s; p_i) = \begin{cases} p_i, & \text{if } X_s = 1 \\ 1 - p_i, & \text{if } X_s = 0 \end{cases},$$

where $p_i = \Pr(X_s = 1) = \Pr(c_s \le r_i)$.

Second, we analyze the participation behavior of a crowd of users in a grid. If the population in grid i is known as n_i, the number of participants, Y_i, can be represented as $Y_i = \sum_{s=1}^{n_i} X_s$. Note that Y_i is a function of n_i and p_i. We assume the costs of a crowd of users follow a certain probabilistic model according to the law of large numbers, which can be learned given the numbers of participants under different incentives. Take the uniform distribution as an example, e.g., $c_s \sim \mathcal{U}(c_{min}, c_{max})$, where c_{min} and c_{max} represent the lower and upper bound respectively. We can deduce that given a reward r_i, Y_i obeys the Binomial distribution as $Y_i \sim B(n_i, p_i)$, where $p_i = \frac{r_i - c_{min}}{c_{max} - c_{min}}$. Therefore, the probability of collecting γ_i measurements equals

$$\Pr(Y_i = \gamma_i) = \frac{n_i!}{\gamma_i!(n_i - \gamma_i)!} p_i^{\gamma_i} (1 - p_i)^{n_i - \gamma_i}.$$

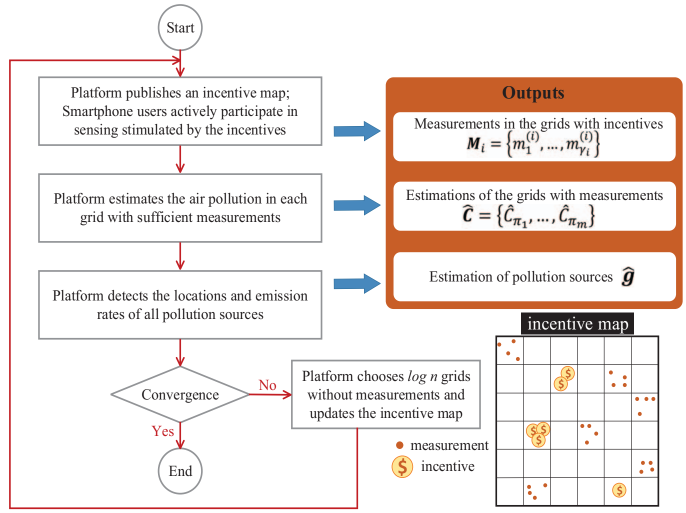

Fig. 1. The workflow of our iterative scheme and the output of each step in an iteration.

III. Scheme for Adaptive Incentive Design

In this section, we propose an iterative scheme for the crowdsensing process and describe how to design incentives adaptively in each iteration.

A. Overview

The workflow of our proposed iterative scheme contains four major steps, as illustrated in Fig. 1. To implement this scheme, four key issues should be addressed:

- **Reliability/Quality of measurements**. sufficient measurements from the same grid are needed to eliminate their measuring errors cooperatively. However, how many measurements can guarantee the accuracy is unknown.
- **Unknown valuable grids in CS**. Choosing grids to collect measurements which are valuable for recovering the whole pollution map via CS, can reduce the amount of needed measurements and the total payment. However, how to evaluate the value of each grid in CS is unsolved.
- **Number of iterations**. It is difficult to determine the number of iterations (namely judging the convergence), because the error between the estimation of pollution sources and the ground truth cannot be calculated directly.
- **Balance between accuracy and payment**. In terms of choosing grids to collect measurements, the needs in detecting pollution sources and the rewards required to collect enough measurements should be balanced.

In response to these issues, we describe each step in detail in the following subsections respectively.

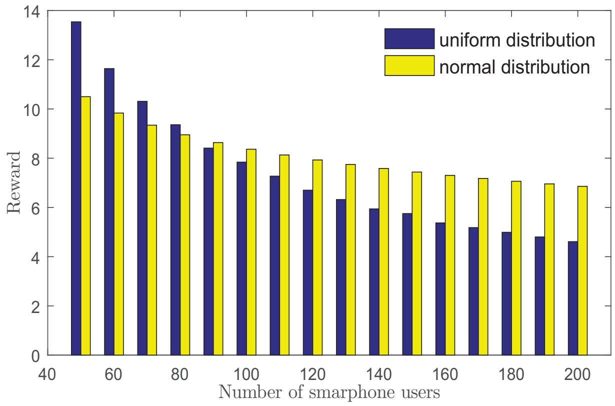

Fig. 2. Reward set for each measurement v.s. number of smartphone users.

B. Pollutant concentration estimation

At the beginning of an iteration, smartphone users in grids with incentives are stimulated to participate in sensing. Based on the measurements collected from one grid, the platform can estimate the real pollution concentration of this grid. To achieve a high accuracy of the estimation, a certain number (denoted by Υ, e.g., 30) of measurements are needed due to the measuring errors of smartphones.

To ensure at least Υ measurements can be collected in grid i with a high probability, e.g., 99.5%, a proper reward should be set for the grid, which can be deduced according to the participation model as follows

$$1 - \sum_{\gamma_i=0}^{\Upsilon-1} \Pr(Y_i = \gamma_i) \geq 99.5\%.$$

Thus, p_i can be computed by solving the above inequality. Accordingly, reward r_i can be calculated. Fig. 2 plots the proper reward r_i when varying the number of smartphone users n_i under two different cost distributions with parameters $c_{min} = 1$ and $c_{max} = 20$. We can observe that as the number of smartphone users grows, the required reward declines, while the marginal decrease becomes smaller and smaller.

C. Pollution source detection

Next, we show how to employ CS in pollution source detection based on the estimated pollutant concentrations $\widehat{\mathbf{C}}$ in sampled grids. As we have $\mathbf{C} = \mathbf{\Omega}\mathbf{g}$, $\mathbf{C}$ can be seen as decomposed into $\mathbf{g}$ based on basis vectors $\mathbf{\Omega}$, although $\mathbf{\Omega}$ is not orthogonal. Due to the sparsity of $\mathbf{g}$, CS can be employed to recover the whole pollution map based on sampled concentration estimations $\widehat{\mathbf{C}}$. We denote the vector of sampled concentration estimates as $\widehat{\mathbf{C}} = \{\widehat{C}_{\pi_1}, \cdots, \widehat{C}_{\pi_m}\}^T$, where $\pi_i \in \{1, \cdots, n\}, 1 \leq i \leq m$. Here, $\boldsymbol{\pi} = \{\pi_1, \cdots, \pi_m\}$ marks the grids have estimated pollution concentrations already. The corresponding transaction matrix is $\mathbf{\Omega}' = \{\Omega_{\pi_i}\}_{i=1}^{m}$. According to CS, $\mathbf{g}$ can be estimated as $\widehat{\mathbf{g}}$ by solving the following problem,

$$\arg\min_{\widehat{\mathbf{g}}} \|\widehat{\mathbf{g}}\|_{\ell_1}, \text{ s.t. } \widehat{\mathbf{C}} = \mathbf{\Omega}'\widehat{\mathbf{g}}. \tag{2}$$

Algorithm 1: Grid selection for updating incentive map

Input: number of grids n, transition matrix $\mathbf{\Omega}$, reward matrix $\mathbf{R}$, sampled grids $\boldsymbol{\pi}^{(t)}$ and the estimate of pollution sources $\widehat{\mathbf{g}}^{(t)}$ and $\widehat{\mathbf{g}}^{(t-1)}$
Output: Updated sampled grids $\boldsymbol{\pi}^{(t+1)}$
`// Initialization`
$\boldsymbol{\pi}^{(t+1)} = \boldsymbol{\pi}^{(t)}$ and $\widehat{\mathbf{g}}_0^{(t+1)} = \widehat{\mathbf{g}}^{(t)}$;
$\mathbf{V} = \mathbf{\Omega} * |\widehat{\mathbf{g}}^{(t)} - \widehat{\mathbf{g}}^{(t-1)}|$;
`// Choose grids one by one`
for $j \leftarrow 1$ **to** $\lfloor \log n \rfloor$ **do**
  $\mathbf{VRR} = \frac{\mathbf{V}}{\mathbf{R}}$;
  `// Choose a new grid`
  $\pi_j = \arg\max_{i \notin \boldsymbol{\pi}^{(t+1)}} VRR_i$;
  $\boldsymbol{\pi}^{(t+1)} = \boldsymbol{\pi}^{(t+1)} \bigcup \{\pi_j\}$;
  `// Update the values of grids`
  Solve Problem (2) and obtain estimate $\widehat{\mathbf{g}}_j^{(t+1)}$ with
  $$\widehat{\mathbf{C}} = \{\widehat{C}_{\boldsymbol{\pi}^{(t+1)}}, \Omega_{\pi_j} * \widehat{\mathbf{g}}_{j-1}^{(t+1)}\}^T,$$
  $$\mathbf{\Omega}' = \{\Omega_{\pi_i}\}, \text{ where } \pi_i \in \boldsymbol{\pi}^{(t+1)}$$
  $$\mathbf{V} = \mathbf{\Omega} * |\widehat{\mathbf{g}}_j^{(t+1)} - \widehat{\mathbf{g}}_{j-1}^{(t+1)}|;$$
end
Return $\boldsymbol{\pi}^{(t+1)}$;

However, due to the unknown $\mathbf{g}$, it is non-trivial to judge when the convergence is achieved. To test the difference between our estimate $\widehat{\mathbf{g}}$ and ground truth $\mathbf{g}$, we adopt the *k-fold cross-validation*. Cross-validation is a technique for assessing how accurately a predictive model will perform in practice given a set of labeled samples. In a round of cross-validation, the labeled dataset is divided into two subsets: one is used to train the model (training set); the other is used to validate (testing set). The mean squared error (MSE) of the testing set is always used to assess the accuracy of the trained model. Multiple rounds are performed using different partitions to reduce the assessment error. In our scheme, we use 5-fold cross validation, which means concentration estimations $\widehat{\mathbf{C}}$ are randomly partitioned into five equal sized subsets. In each round, four subsets compose the training set, used to estimate $\mathbf{g}$ by solving (2) while the other subset is used to test the difference between $\widehat{\mathbf{g}}$ and $\mathbf{g}$.

D. Incentive map update

As shown in Fig. 1, updating the incentive map in each iteration is converted into choosing $\lfloor \log n \rfloor$ new grids if the estimate of pollution sources does not converge. The reward set for each chosen grid is decided by its population as mentioned in Section III-B. With the measurements collected in the chosen grids, the platform aims at enhancing the accuracy of pollution source detection and maintaining a low payment. Therefore, for each grid, two aspects need to be considered for deciding whether it should be chosen: 1) the rewards paid for collecting enough measurements, and 2) the marginal value in detecting pollution sources via CS.

First, we define a mathematical metric to measure the marginal value of each grid, given the current sampled grids. We denote the estimates of pollution sources in iteration $(t-1)$

and iteration (t) as $\widehat{\mathbf{g}}^{(t-1)}$ and $\widehat{\mathbf{g}}^{(t)}$. For an arbitrary grid $i \in \{1, \cdots, n\}$, its value v_i is defined as the *difference* between the two pollutant concentrations computed based on $\widehat{\mathbf{g}}^{(t-1)}$ and $\widehat{\mathbf{g}}^{(t)}$ respectively. Thus, we have

$$\mathbf{V} = \mathbf{\Omega} * |\widehat{\mathbf{g}}^{(t)} - \widehat{\mathbf{g}}^{(t-1)}|, \tag{3}$$

where $\mathbf{V} = \{v_1, \cdots, v_n\}$.

Then, we design an algorithm for selecting $\lfloor \log n \rfloor$ grids by balancing their values and rewards at the same time. As shown in Algorithm 1, grids are selected one by one. In each time, grid π_j with the maximum Value-to-Reward-Ratio (VRR) is picked up, where VRR of each grid is computed as the ratio of its value to its reward (line 4). Then, marginal values need to be updated (line 7) because estimate $\widehat{\mathbf{g}}$ changes after the pollutant concentration of a new grid is added into $\widehat{\mathbf{C}}$. Note that the pollutant concentration of the selected grid cannot be estimated because measurements are not collected until the next iteration. Thus, we use the value of $\Omega_{\pi_j}\widehat{\mathbf{g}}$ as a replacement.

IV. Performance Evaluation

In this section, we evaluate the performance of our proposed scheme, especially Algorithm 1.

A. Methodology and setups

It has been proven by many previous works [15] that compressive sensing performs well in recovering a sparse vector from a few samples, compared with common interpolation methods, such as linear interpolation and Kriging interpolation. Thus, in our simulations, we concentrate on showing the performance of our proposed grid selection algorithm, compared with three baseline algorithms:

- Random. This algorithm randomly chooses $\lfloor \log n \rfloor$ new grids in each iteration until convergence.
- Greedy-Reward. The cheapest $\lfloor \log n \rfloor$ grids are selected in each iteration until convergence.
- Greedy-Value: In each iteration, this algorithm picks up grids one by one according to their values calculated as (3), which are updated as well as Algorithm 1.

Three metrics are used to evaluate the performance of the four algorithms from different aspects: 1) **Error of estimation**. This metric is calculated as $\frac{\|\widehat{\mathbf{g}}-\mathbf{g}\|_{\ell_2}}{n}$, where $\widehat{\mathbf{g}}$ is estimated according to (2) based on all measurements collected until convergence. It measures the accuracy of detecting pollution sources based on our iterative scheme; 2) **Total payment**. We sum up the rewards of all selected grids as $\sum_{i=\pi_1}^{\pi_m} r_i$ to represent the total payment, assuming the amount of measurements collected in each sampled grid is equal; 3) **Number of iterations**. This metric is proportional to the total number of sampled grids as well as the time consumed for the crowdsensing process.

The default setting of system parameters is as follows. All simulations are conducted on a square area divided into $50*50$ grids ($n = 2500$), and the size of each grid is equal to $200\text{m} * 200\text{m}$. The wind blows at $\nu = 5\text{m/s}$. Transition matrix $\mathbf{\Omega} \in \mathbb{R}^{2500*2500}$ can be computed according to (1) given $\alpha = 200$ and $\beta = 1000$. The locations and emission rates of pollution sources are randomly chosen from $\{1, \cdots, 2500\}$ and $\{1000\text{mg/s}, \cdots, 5000\text{mg/s}\}$, respectively. The population n_i in each grid is randomly generated varying in $[50, 200]$. We conduct simulations considering both uniform distribution and normal distribution for the costs of smartphone users, with $c_{min} = 1$ and $c_{max} = 20$. Given the value of n_i, the value of reward r_i can be known according to Fig. 2. We study the performance of the four algorithms by varying the number of pollution sources k from 5 to 25. Each setting has ten runs.

B. Simulation results

Fig. 3$\sim$5 plot the performance of the four algorithms under uniformly distributed costs, while Fig. 6$\sim$8 plot the performance under normal distributed costs.

As shown in Fig. 3 and Fig. 6, pollution sources can be accurately detected via compressive sensing based on sufficient measurements. The result demonstrates that the cross-validation method works well for judging the convergence of our iterative scheme. We can find that Greedy-Value algorithm and Algorithm 1 perform better in different settings, compared with other two algorithms. Moreover, all algorithms perform better under uniform distributed costs than normal distributed costs because that more measurements are collected as shown in Fig. 5 and Fig. 8.

Fig. 4 and Fig. 7 show that more payment is needed as k increases. Although Greedy-Reward algorithm chooses the cheapest grids in each iteration, it consumes more money compared with Greedy-Value algorithm and Algorithm 1. This is because the cheapest measurements may suffer poor values in pollution source detection, which leads to collecting more measurements and thus incurring a high payment. From Fig. 5 and Fig. 8, we can find that the amounts of measurements needed by Greedy-Value algorithm and Algorithm 1 are approximately equal. However, the total payment achieved by Algorithm 1 is much less than Greedy-Value algorithm in different settings. When there are 20 pollution sources, Algorithm 1 can save 53.43% and 22.92% payment compared with Greedy-Value algorithm under two different cost distributions.

V. Related Work

A series of studies [8][9][10] dedicate to designing incentive mechanisms for crowdsensing applications to stimulate smartphone users participating in sensing. In [8], a recurrent reverse auction is employed to select participants according to their locations given constraints in budget and coverage. Koutsopoulos [9] derives a mechanism, which minimizes the total cost paid to participants by tracking the quality of their reported cost information and using it for determining participation level and payment. In [10], Zhao *et al.* propose online incentive mechanisms by considering smartphone users randomly arriving one by one. Under a budget, the value of services provided by participants is maximized before a given deadline. All these mechanisms collect private information from each smartphone user before sensing process, and focus on making them truthful. Different from these studies, we analyze the participation model in terms of a group of users rather than an individual, which

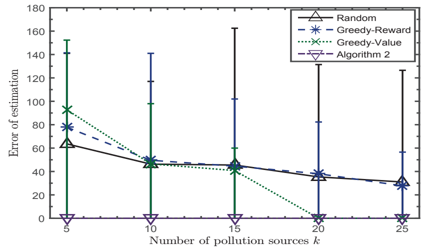

Fig. 3. Error of estimation vs. number of pollution sources under uniform distributed costs.

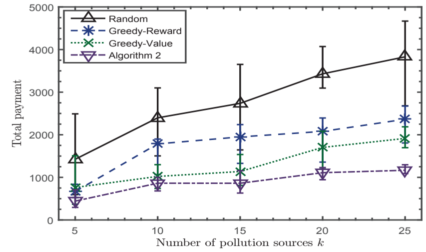

Fig. 4. Total payment vs. number of pollution sources under uniform distributed costs.

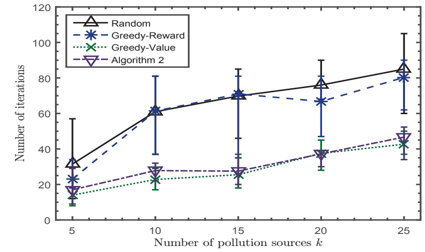

Fig. 5. Number of iterations vs. number of pollution sources under uniform distributed costs.

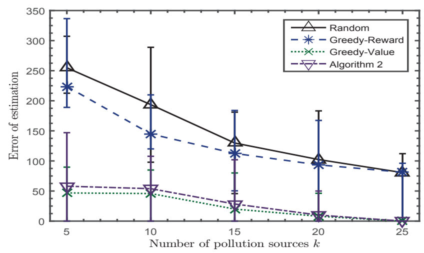

Fig. 6. Error of estimation vs. number of pollution sources under normal distributed costs.

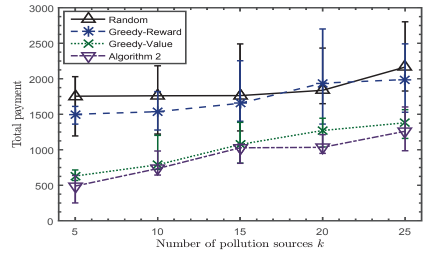

Fig. 7. Total payment vs. number of pollution sources under normal distributed costs.

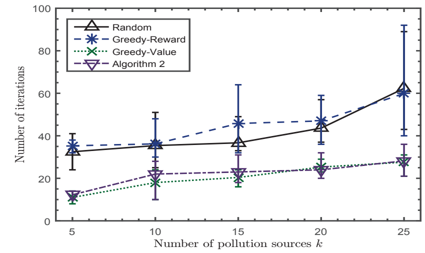

Fig. 8. Number of iterations vs. number of pollution sources under normal distributed costs.

follows statistic laws. Therefore, incentives can be designed according to the population in interested areas.

Compressive sensing has been proved to be efficient in reducing the amount of sampled data and few works [13][14][15] have applied it in crowdsensing.

VI. Conclusion

This paper has focused on the problem of minimizing the total payment for accurately detecting air pollution sources based on crowdsensing via designing an efficient incentive mechanism to stimulate smartphone user participation. We have proposed an iterative scheme for the sensing process, where the platform updates the incentive map iteratively according to the current collected measurements. Moreover, we have shown how to apply the EM algorithm in estimating pollutant concentration in a grid based on measurements, and how to employ CS in recovering the whole pollution map without collecting measurements everywhere. Comprehensive simulations have been conducted to confirm the superiority of our proposed algorithms.

Acknowledgments

This research was supported in part by China 863 Program (No. 2015AA015303), NSFC (No. 61472254, 61170238), the Science and Technology Commission of Shanghai (Grant No.14511107500 and 15DZ1100305), Program for New Century Excellent Talents in University, and Singapore NRF (CREATE E2S2). The research is also partially supported by the U.S. National Science Foundation under Grant No. CSR-1513719.

References

[1] Y. Zheng, F. Liu, and H.-P. Hsieh, "U-air: When urban air quality inference meets big data," in *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2013, pp. 1436–1444.

[2] https://en.wikipedia.org/wiki/Air_Quality_Egg.

[3] https://www.rmit.edu.au/news/all-news/media-releases/2015/october/revolutionary-new-weapon-in-air-pollution-fight/.

[4] P. Dutta, P. M. Aoki, N. Kumar, A. Mainwaring, C. Myers, W. Willett, and A. Woodruff, "Common sense: participatory urban sensing using a network of handheld air quality monitors," in *Proceedings of the 7th ACM conference on embedded networked sensor systems*. ACM, 2009, pp. 349–350.

[5] D. Mendez, A. J. Perez, M. Labrador, J. J. Marron *et al.*, "P-sense: A participatory sensing system for air pollution monitoring and control," in *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*. IEEE, 2011, pp. 344–347.

[6] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," *Mobile Sensing*, 2012.

[7] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 173–184.

[8] L. G. Jaimes, I. Vergara-Laurens, M. Labrador *et al.*, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*. IEEE, 2012, pp. 103–108.

[9] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Infocom, 2013 proceedings ieee*. IEEE, 2013, pp. 1402–1410.

[10] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *INFOCOM, 2014 Proceedings IEEE*. IEEE, 2014, pp. 1213–1221.

[11] M. R. Beychok, "Fundamentals of stack gas dispersion," 1995.

[12] D. L. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[13] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "Cost-aware compressive sensing for networked sensing systems," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 130–141.

[14] L. Wang, D. Zhang, A. Pathak, C. Chen, H. Xiong, D. Yang, and Y. Wang, "Ccs-ta: quality-guaranteed online task allocation in compressive crowdsensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 683–694.

[15] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "More with less: lowering user burden in mobile crowdsourcing through compressive sensing," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015, pp. 659–670.

[16] https://en.wikipedia.org/wiki/Outline_of_air_pollution_dispersion.