# *TRAC*: Truthful Auction for Location-Aware Collaborative Sensing in Mobile Crowdsourcing

Zhenni Feng[†,♮], Yanmin Zhu[†,♮,*], Qian Zhang[§], Lionel M. Ni[†,§], Athanasios V. Vasilakos[‡]
[♮] Shanghai Key Lab of Scalable Computing and Systems; [*] Corresponding Author
[†] Department of Computer Science and Engineering, Shanghai Jiao Tong University
[§] Hong Kong University of Science and Technology
[‡] National Technical University of Athens
[†] {zhennifeng, yzhu}@sjtu.edu.cn; [§] {qianzh, ni}@cse.ust.hk; [‡] vasilako@ath.forthnet.gr

***Abstract*—In this paper, we tackle the problem of stimulating smartphone users to join mobile crowdsourcing applications with smartphones. Different from existing work of mechanism design, we uniquely take into consideration the *crucial dimension of location information* when assigning sensing tasks to smartphones. However, the location awareness largely increases the theoretical and computational complexity. In this paper, we introduce a *reverse auction framework* to model the interactions between the platform and the smartphones. We rigorously prove that optimally determining the winning bids is *NP hard*. In this paper we design a mechanism called *TRAC* which consists of two main components. The first component is a near-optimal approximate algorithm for determining the winning bids with polynomial-time computation complexity, which approximates the optimal solution within a factor of $1 + \ln(n)$, where $n$ is the maximum number of sensing tasks that a smartphone can accommodate. The second component is a critical payment scheme which, despite the approximation of determining winning bids, guarantees that submitted bids of smartphones reflect their real costs of performing sensing tasks. Through both rigid theoretical analysis and extensive simulations, we demonstrate that the proposed mechanism achieves truthfulness, individual rationality and high computation efficiency.**

## I. Introduction

With the rapid advance in mobile devices and their embedded sensors, *mobile crowdsourcing with smartphones* has been a compelling paradigm for collecting distributed sensory data to share with the general public [1]. Being carried by human users who locate in different places, smartphones can easily collect ubiquitous data and share such data with a large number of potential users [2] [3].

A mobile crowdsourcing system typically consists of a platform residing on the cloud and mobile smartphones. An example is illustrated in Fig. 1. The platform selects the set of smartphones to jointly provide sensing services. Once receiving the notification of the platform, a smartphone collects the required data. It then returns the collected sensing data to the platform which then forwards the data to the platform user. This demonstrates that a mobile crowdsourcing system with geographically distributed smartphones can support a wide range of large-scale monitoring applications [4].

*Stimulating smartphone users to join collaborative sensing is greatly important to the success of mobile crowdsourcing with smartphones*. As we know, it incurs some cost (*e.g.,* power consumption for driving sensors and the CPU) when a smartphone performs sensing for others. In addition, smartphone users may suffer privacy breach as they share sensed data (*e.g.*, location data) with others. Thus, smartphone users may not be willing to join a mobile crowdsourcing system unless they receive enough compensation for their resource consumption. Without the participation of a large number of smartphones, one is not able to collect the desired data through the mobile crowdsourcing application. As a result, no real mobile crowdsourcing applications would be widely adopted. Although a number of mobile crowdsourcing applications [5] [6] [7] [8] have been proposed, most of them have assumed that smartphones voluntarily contribute to the mobile crowdsourcing system, which is not true in the real world.

The problem of stimulating smartphone users to join mobile crowdsourcing applications is highly complicated because *strategic* behaviors of smartphone users can seriously hinder the potential collaboration of smartphone users. Merely providing fixed monetary rewards to participating smartphones, however, is not adequate for guaranteeing smartphones to perform collaborative sensing. Strategic smartphones would take actions solely to maximize their own payoffs. In addition, the cost information of each smartphone is *private* and no smartphones would release such information. As a result, the platform has no access to the cost information of each smartphone. To maximize its own payoff, a smartphone may

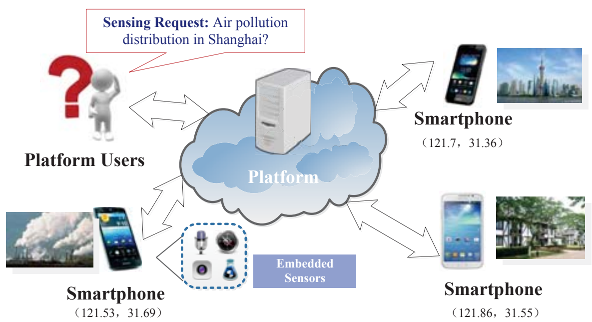


Fig. 1. An example of mobile crowdsourcing applications. Smartphones are distributed over a large city. The location of each smartphone is described by a 2-tuple, $(longitude, latitude)$. A platform user advertises sensing tasks of collecting air pollution data through the platform residing on the cloud, and smartphones can contribute to the sensing tasks by returning their sensed data (*e.g.*, a photo of its surroundings) to the platform.

deliberately claim a cost higher than the real one. Thus, mobile crowdsourcing applications should be resilient to strategic behaviors of smartphones.

There have been several research efforts on developing incentive mechanisms for mobile crowdsourcing applications, which can generally be divided into two categories. One category of existing work tries to exploit auctions for inducing cooperation from smartphones. In [2], Yang *et al.* consider two types of incentive mechanisms for maximizing the utility of the platform. It first proposes one incentive mechanism using the Stackelberg game. This mechanism assumes that the utility model of each user is known to the platform, which is not always true. It then proposes an auction-based mechanism. In this mechanism, a smartphone can only bid for a specific subset of sensing tasks, which is not realistic in the real world. The second category of existing work [9] [10] adopts a different approach to stimulating cooperative behaviors of selfish smartphone users. They try to build trust among smartphones. Trust is useful but may not be sufficient to fighting against strategic behaviors of smartphone users.

*We also have an important observation that location awareness is central to most mobile crowdsourcing applications*. A sensing task typically specifies the location where the sensing task should be performed. This is because that the desired sensing data are closely related to the specific location. A sensing data collected at an irrelevant location is meaningless or even invalid. This practical consideration on location awareness caters to *more meaning and accurate matching between demands and supplies of sensing services*. Unfortunately, most of existing designs of incentive mechanism [11] [12] have neglected this important dimension of location information in their designs.

In this paper, we introduce a practical *reverse auction framework*, in which the platform announces sensing tasks each of which has a location attribute, and smartphones can submit bids for a set of tasks within their service coverage. Each bid has a *private* cost for performing the set of tasks. The platform then decides the set of winning bids by which the sensing tasks are allocated to the corresponding smartphones. The platform will pay an amount of rewards to each contributing smartphone for their sensing services.

To make this reverse auction framework actually work for location-aware collaborative sensing in mobile crowdsourcing, we aim at designing a truthful mechanism by which each smartphone would truthfully declare its cost. To solve the combinatorial auction problem, we have to address *two critical problems*: 1) we have to design an efficient algorithm to determine the winning bids and allocate the sensing tasks to smartphones with a high efficiency (minimal social cost), and 2) we have to design a payment policy to guarantee truthfulness, fighting against the strategic behavior of smartphones. However, we find that, after rigorous proof, the optimal winning bids determination problem is *NP hard*. In this paper we design a truthful mechanism called *TRAC* which consists of *two main components*. Exploiting an efficient approximate algorithm, the first component determines the winning bids. The second component determines the critical payment to each winning bid. The approximation algorithm can approximate the optimal solution within a factor of $1 + \ln(n)$, with polynomial-time computation complexity, where $n$ is the maximum number of sensing tasks that a smartphone can accommodate. We also theoretically prove that despite the approximate assignment algorithm, *TRAC* guarantees truthfulness.

We highlight the main intellectual contributions as follows.

- We consider location awareness in the design of a truthful incentive mechanism for location-aware collaborative sensing in mobile crowdsourcing. However, the consideration on location information essentially increases the problem complexity of combinatorial auction design.
- We design an algorithmic mechanism for inducing location-aware collaborative sensing in mobile crowdsourcing. We prove that optimally determining the winning bids with location awareness is *NP hard*. The proposed mechanism consists of a *polynomial time and near-optimal task allocation algorithm* and a novel payment scheme that guarantees the truthfulness of the proposed mechanism.
- Through both rigid theoretical analysis and extensive simulations with real trace data sets, we demonstrate that the proposed mechanism achieves the desired properties of truthfulness, individual rationality and high computation efficiency. In addition, the proposed algorithms achieve low social cost and modest overpayment.

The remainder of the paper is organized as follows. In Section II, we first present the system model, the reverse auction framework and the mathematical problem formulation. Then, the detailed design of *TRAC* is described in Section III. The theoretical analysis of the proposed mechanism is presented in Section IV. Then, we review related work in Section VI. Finally, we conclude the paper and discuss future research directions in Section VII.

## II. System Model and Problem Formulation

### A. System Model of Location-aware Mobile Crowdsourcing

We consider a mobile crowdsourcing system with smartphones consisting of a *platform* and many *smartphone users*. The platform resides in the cloud. The platform accepts sensing requests from *platform users* who connect to the platform via the cloud. The platform periodically publicizes sensing tasks to be performed by smartphones. Let $T$ denote the set of sensing tasks, $T = \{t_1, t_2, \cdots, t_m\}$. A sensing task $t_i$ specifies the desired sensing service and the corresponding location where the sensing data should be collected. Let $p(t_i)$ denote the location of the sensing task. A sensing task is atomic, meaning that it is either entirely performed or it is not completed.

There are $n$ smartphones which are interested in performing sensing tasks and the set of smartphones is denoted by $N = \{1, 2, \cdots, n\}$. Each smartphone $i$ is aware of its own location $l_i$, through Global Positioning Systems (GPS) or other localization schemes [8]. Each smartphone $i$ is intrinsically

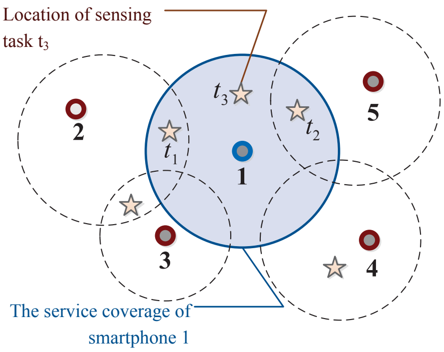


Fig. 2. An example of service coverage of smartphones. The blue area denotes the service coverage of smartphone 1. Three sensing requests fall within the coverage of smartphone 1. Thus, smartphone 1 can provide sensing service to each of the three tasks. The sensing coverage of different smartphones can be different.

associated with a geographical service coverage, denoted by $\vartheta_i$ (as illustrated in Fig. 2). Only those sensing tasks within the service coverage may potentially receive sensing service of smartphone $i$.

The service coverage can be different from smartphone to smartphone, and is dependent on the smartphone user and the associated preferences or restrictions. It is practical to assume that the service coverage $\vartheta_i$ of smartphone $i$ is dependent on the current location $l_i$ of the smartphone. Thus, given the current location $l_i$ of smartphone $i$, one is able to determine its service coverage $\vartheta_i$. It is worth noting that each smartphone can have a different function mapping its current location to its service coverage $\vartheta_i$. We assume that each smartphone in the system would share such information with the platform. With $\vartheta_i$, the platform is able to determine the subset $T_i$ of sensing tasks, $T_i \subseteq T$, that smartphone $i$ is able to provide sensing service. Each smartphone should not misreport its own service coverage. Misreporting may result in failure of completing a sensing task and a serious penalty would be reinforced. Protection of location privacy of smartphones and is beyond the scope of this paper and subject to future research.

### B. Reverse Auction Framework

In the mobile crowdsourcing system, smartphones compete for opportunities to provide sensing services. We introduce a *reverse auction framework* for modeling the interactions between the platform and the smartphones, in which smartphones are the *sellers* and the platform is the *buyer* (buying sensing services). The framework is such called as it is a type of auction in which the roles of buyer and seller are reversed.

With the framework, the interactions between the platform and smartphones are described as follows, which is also illustrated in Fig. 3.

1) The platform advertises a set $T$ of sensing tasks to all the smartphones in the mobile crowdsourcing system.
2) Each user $i$ replies with a set $B_i$ of $k_i$ bids, each of which is a tasks-bid pair $\beta_i^k = (Q_i^k, b_i^k)$, in which $Q_i^k$ is a subset of sensing tasks that are within its geographical service coverage, $Q_i^k \subset T_i$, and $b_i^k$ is called *claimed cost* of the subset of task $Q_i^k$ which is *reserved price* that user $i$ wants to charge for the service. In addition, each user $i$ submits a number $r_i$ which is the maximum number of bids that it can accommodate. Note that a smartphone user could not submits two bids for the same set of sensing tasks.
3) The platform determines whether a bid is winning or not, *i.e.*, it selects a subset $S$ from all submitted bids, $S \subseteq \bigcup_{i\in N} B_i$, in which $\beta_i^k \in S$ indicates that smartphone $i$ would perform the set of sensing tasks $Q_i^k$ in its bid $\beta_i^k$.
4) Each smartphone $i$ performs the sensing tasks in its winning bids which are $S \bigcap B_i$ and sends the sensed data back to the platform.
5) Each smartphone $i$ is paid an amount of money $p_i^k$ for its winning bid $\beta_i^k$, for each $\beta_i^k \in S \bigcap B_i$.

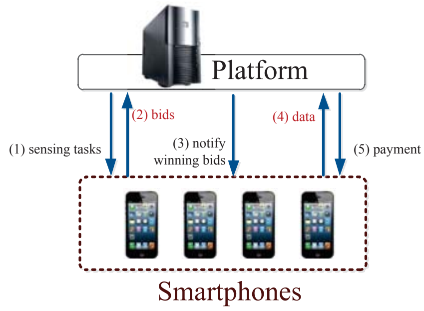


Fig. 3. The interactions between the platform and smartphones within the reverse auction model.

Due to the distributed nature of mobile crowdsourcing, the real cost $c_i^k$ of performing the set $Q_i^k$ of the sensing tasks is *private* and *unknown* to others. Each smartphone is owned by a selfish individual who always tries to maximize its own payoff. Thus, it is possible for a smartphone to manipulate the claimed cost for its own good, *i.e.*, $b_i^k$ may not be equal to $c_i^k$. This kind of behavior is typical *strategic behavior*. The existence of strategic behaviors makes it difficult for the platform to *hire* those smartphones with lower costs.

Smartphones are strategic and hence each smartphone $i$ claims cost $b_i^k$ for bid $\beta_i^k$ that may be different from the real cost $c_i^k$ for maximizing its own payoff (or benefit). We define the payoff of a smartphone as follows.

**Definition 1 (Smartphone Payoff).** *The payoff of a smartphone is the sum of payoffs of all its winning bids. The payoff of a winning bid is the difference between the payment it receives and its real cost. The payoff of the smartphone can be computed as follows,*

$$u_i = \sum_{\beta_i^k \in S \bigcap B_i} (p_i^k - c_i^k). \tag{1}$$

### C. Problem Formulation

We next give the mathematical formulation of the mechanism design problem. In this work, the platform determines the winning bids and allocates sensing tasks to the corresponding smartphones.

**Definition 2** (**Winning Bids Determination Problem (WBDP)**). *For each subset set $S$ of bids, the winning bids determination problem is defined as follows:*

$$min \sum_{\beta_i^k \in S} c_i^k \quad (2)$$

$$s.t.\ |S \bigcap B_i| \leq r_i \quad (3)$$

$$\bigcup_{\beta_i^k \in S} Q_i^k = T. \quad (4)$$

*Remarks*: The definition of WBDP shows the objective of the platform selecting the winning bids is to minimize the *social cost* which is the sum of the real costs of smartphones completing all the sensing tasks. The first constraint indicates that the number of winning bids of user $i$ could not exceed its maximum number $r_i$. The second constraint shows that the platform guarantees that each sensing task is finished unless a sensing task is within less than one submitted bids. Notice that if no bid that covers a sensing task $t_j$ and then it is obvious that $t_j$ could not be finished. And we also exclude the situation where only one bid covers the sensing task $t_j$ in order to prevent the monopoly. Thus, we assume that there are enough smartphones within the system and more than one smartphones compete for each task.

**Definition 3** (**Truthful Mechanism Design Problem (TMDP)**). *For each bid of smartphone i, let $\tilde{\beta_i^k} = (Q_i^k, c_i^k)$ denote the truthful bid and $\beta_i^k = (Q_i^k, b_i^k)$ denote the untruthful bid. The payoffs of user $i$ for the truthful bid and the untruthful bid is $u_i(\tilde{\beta_i^k}) = p_i(\tilde{\beta_i^k}) - c_i^k$ and $u_i(\beta_i^k) = p_i(\beta_i^k) - c_i^k$, respectively. The TMDP problem is to design a* payment scheme *such that*

$$u_i(\tilde{\beta_i^k}) \geq u_i(\beta_i^k). \quad (5)$$

*Remarks*: A payment scheme resulting from the TMDP can guarantee that smartphones declare their costs truthfully.

The goal of our work is to design a truthful mechanism that solves the two problems defined above. This mechanism should also have the following desired properties.

**Definition 4** (**Individual Rationality**). *The payoff of each bid of user $i$ is nonnegative, $p_i^k \geq c_i^k$.*

*Remarks*: To stimulate smartphones to participate in the mobile crowdsourcing applications, a basic condition is that the cost of each smartphone must be covered.

**Definition 5** (**Computational Efficiency**). *An algorithm has the property of computational efficiency if it terminates in polynomial time.*

*Remarks*: The computational efficiency of the algorithms of solving WBDP and TMDP is of great importance in realistic scenarios. Any optimal algorithm with high complexity is useless in reality.

## III. Design of *TRAC*

*TRAC* consists of two components: The first component solves WBDP to determine the winning bids, and the second component is a payment scheme for solving TMDP. Before describing the algorithm for the first component, we first analyze the complexity of solving the WBDP, *i.e.*, determining the winning bids to minimize the social cost. We regorously prove that WBDP is NP hard. Then, we propose an algorithm that obtains a near-optimal solution with low computational complexity, which is different from traditional truthful mechanism. Finally, we propose the payment scheme for solving TMDP to induce smartphones to disclose their real costs truthfully.

### A. Complexity Analysis of WBDP

It is very important to solve WBDP with a time efficient algorithm. Unfortunately, as we are going to prove next, WBDP is NP hard.

**Theorem 1.** *The WBDP is NP hard.*

*Proof:* To prove WBDP is NP-hard, we can prove that its decision version is NP-complete. For the decision problem, we should demonstrate that it belongs to NP, and then find another known NP-complete problem that could be reduced to the decision version of WBDP in polynomial time.

The decision version of our problem is a modified minimum weighted set cover (MWSC) problem in which some pairs of subsets are mutually exclusive, *i.e.,* they could not win simultaneously. The decision problem belongs to NP as checking whether a solution is correct or not could end up in polynomial time.

Next, we use the decision version of minimized weighted set cover as the known NP-complete problem. An instance of the known problem is defined as follows.

**Definition 6** (**An Instance of MWSC**). *For a universe set $U$ and a set $S = \{s_1, s_2, \cdots, s_n\}$, each $s_i$ satisfies $s_i \subseteq U$ and its weight is $w(s_i)$. The question is whether exists a set $Q \subseteq S$ and $\sum_{s_i \in Q} w(s_i) \leq K$. We regard the instance as $A$ in later discussion.*

Next, we change the instance of MWSC to an instance of our problem. We construct a set $\tilde{S} = \{s_1, s_2, \cdots, s_n, z\}, z \subset U$, $w(z) > K$ and $z$ could not be chosen together with each $s_i$. Thus, we get an instance of our problem, which is denoted by $B$.

Then, we can simply see that $q$ is a solution of $A$ if and only if $q$ is a solution of $B$. Moreover, the reduction from $A$ to $B$ ends in polynomial time. ■

*Remarks*: The previous theorem shows that WBDP is NP hard even if each smartphone honestly declares their costs. This essentially rule out the possibility of exploiting the traditionally VCG mechanism [13] for our problem. The VCG mechanism requires that the optimal set of winning bids must be selected, which is impossible since there is no computationally efficient algorithms for solving NP hard

**Algorithm 1: Approximate Algorithm for WBDP**

**Input:** set $T = \{t_1, t_2, \cdots, t_m\}$ of sensing tasks, set $B = \bigcup_{i \in N} B_i$ of all submitted bids, maximum number $r_i$ of winning bids for user $i$.
**Output:** set $S$ of winning bids, social cost $\omega$.
// Initialization
$S \leftarrow \emptyset,\ \omega \leftarrow 0,\ \tilde{Q} \leftarrow \emptyset$;
**while** $\tilde{Q} \neq T$ **do**
  **for all** $\beta_i^k$ in $B$ **do**
    **if** $Q_i^k \subseteq \tilde{Q}$ **then**
      Remove $\beta_i^k$ from the set $B$;
    **else**
      $r(\beta_i^k) = \frac{c_i^k}{|Q_i^k - \tilde{Q}|}$;
    **end if**
  **end for**
  Sort $r(\beta_i^k)$ for all $\beta_i^k \in B$ in the nondecreasing order and the list is denoted by $\mathcal{R}$;
  //Add a bid into the set of winning bids
  $\beta_s^t$ denotes the head of $\mathcal{R}$;
  $S \leftarrow S \bigcup \beta_s^t,\ \omega \leftarrow \omega + c_s^t,\ \tilde{Q} \leftarrow \tilde{Q} \bigcup Q_s^t$;
  Remove $\beta_s^t$ element from $B$;
  //Delete bids that conflict with $\beta_s^t$
  **for all** $\beta_i^k$ in $B$ **do**
    **if** $\beta_i^k$ conflicts with $\beta_s^t$ **then**
      Delete $\beta_i^k$;
    **end if**
  **end for**
**end while**
**return** $S, \omega$;

problems. Moreover, an approximation algorithm with the VCG mechanism could not guarantee truthfulness.

Consequently, to compose a truthful and computationally efficient mechanism, we have to propose a non-VCG mechanism. We next present the design of an approximate algorithm for solving WBDP.

### B. Approximate Method to Solve WBDP

To achieve the desired property of computational efficiency, we propose an approximate algorithm to solve WBDP. To make it understood easily, we first assume that each smartphone reports its bids truthfully, and then demonstrate that each smartphone would obey the rule of truthfulness in the next section.

The algorithm adopts a greedy strategy to solve the problem. The greedy rule is to pick the next most cost-efficient bid that makes the "greatest progress" towards finishing all sensing tasks until the all sensing tasks are assigned. More specifically, for each bid $\beta_i^k = (Q_i^k, b_i^k)$, its progress towards the goal is the ranking criterion $r(\beta_i^k)$ and is computed as follows:

$$r(\beta_i^k) = \frac{b_i^k}{|Q_i^k - \tilde{Q}|}, \tag{6}$$

where $\tilde{Q}$ denotes the coverage of all winning bids, *i.e.,* $\tilde{Q} = \bigcup Q_t^s$ and $\tilde{Q} = \emptyset$ initially, $\beta_t^s$ denotes each bid that has won before $\beta_i^k$. $|Q_i^k - \tilde{Q}|$ denotes the number of elements each of which belongs to $Q_i^k$ but not $\tilde{Q}$. In each iteration, $\tilde{Q}$ and $r(\beta_i^k)$ is updated and all bids that could not be selected together with existing winning bids are deleted. The pseudo-code is shown in Algorithm 1.

### C. Critical Payment Scheme

The payment to each smartphone should be such determined that it is guaranteed that each smartphone honestly reports its real cost. The rule of *critical payment* introduced in [13] is used to determine the payment to each smartphone.

Each smartphone $i$ is paid an amount of monetary reward for each winning bid $\beta_i^k$. The amount is determined according to a *critical bid* $c(\beta_i^k)$, which is determined as follows: if $\beta_i^k$ satisfies $r(\beta_i^k) > r(c(\beta_i^k))$ bid $\beta_i^k$ wins, while the bid loses if $r(\beta_i^k) < r(c(\beta_i^k))$. The critical bid of each bid $\beta_i^k$ is the first bid that makes $\beta_i^k$ fail. The bid $\beta_i^k$ fails when no progress it could make towards completing all sensing tasks, *i.e.,* $Q_i^k \subseteq \tilde{\mathrm{Q}}$. The payment to $\beta_i^k$ would be related to the claimed cost of its critical bid $c(\beta_i^k)$ and the payment is called critical payment.

The critical bid of a bid $\beta_x^y$ is the first bid $\beta_i^k$ which makes $\beta_x^y$ useless any longer, *i.e.,* all existing winning bids could do all that $\beta_x^y$ could do. The basic idea of finding the critical bid is deleting $\beta_x^y$ and greedily selecting other bids as shown in Algorithm 1 until $\beta_x^y$ is useless ($Q_x^y - \tilde{Q} = \emptyset$). We assume that $\beta_i^k$ is the critical bid and it wins in the $q$-th iteration. Then, the sensing task denoted by bid $\beta_x^y$ is paid an amount of money which is proportional to the ranking criterion of $\beta_i^k$ in that iteration. If we denote the ranking criterion of $\beta_i^k$ in the $q$-th iteration as $r^q(\beta_i^k)$, the critical payment is

$$p_c(\beta_x^y) = |Q_x^y - \tilde{Q}^{q-1}| \cdot r^q(\beta_i^k), \tag{7}$$

where $\tilde{Q}^{q-1}$ denotes the set $\tilde{Q}$ at the end of the $(q-1)-th$ iteration, $|Q_x^y - \tilde{Q}^{q-1}|$ denotes the cardinality of set $Q_x^y - \tilde{Q}^{q-1}$. The algorithm is shown in 2.

Thus, the payoff of each bid is derived as follows.

$$u(\beta_x^y) \ = \ \begin{cases} |Q_x^y - \tilde{Q}^{q-1}| \cdot r^q(\beta_i^k) - c_x^y & \beta_x^y \in S \\ 0 & \beta_x^y \notin S \end{cases}, \tag{8}$$

where $\beta_x^y \notin S, u(\beta_x^y) = 0$ means that losing bids are associated with no payment. The payoff of smartphone $x$ is the sum of all its winning bids, $u_x = \sum_{\beta_x^y \in S} u(\beta_x^y)$.

## IV. Theoretical Analysis

In the section we present theoretical analysis, demonstrating that *TRAC* achieves the desired properties of truthfulness, individual rationality and computational efficiency.

### A. Individual Rationality and Truthfulness

To demonstrate that *TRAC* is truthful, we should reveal that each smartphone will honestly disclose its real costs when the strategies of other smartphones are fixed. According to [13], our proposed mechanism *TRAC* is truthful if and only if the following two conditions hold: (1) the winning bids determination algorithm for WBDP is *monotonic*, and (2) each winning bid is paid the *critical value*.

Before showing *TRAC* satisfies the two conditions, we first define the two conditions of monotonicity and critical value.

**Definition 7** (**Monotonicity**). *For each bid $\beta_i^k$, if $\beta_i^k = (Q_i^k, c_i^k)$ wins, then bid $\tilde{\beta}_i^k$ also wins, $\tilde{\beta}_i^k = (Q_i^k, c_i^k - \delta)$ and $\delta > 0$.*

**Algorithm 2: Critical Payment Scheme for TMDP**

**Input:** bid $\beta_x^y$, other submitted bid $B_{-(x,y)} = \{\beta_s^t | s \neq x \bigwedge t \neq y\}$.
**Output:** critical bid $c(\beta_x^y)$, critical payment $p_c(\beta_x^y)$.
1: $p_c(\beta_x^y) \leftarrow 0,\ \tilde{Q} \leftarrow \emptyset$;
2: **while** $\tilde{Q} \neq T$ **do**
3:    **for all** $\beta_i^k$ in $B_{-(x,y)}$ **do**
4:       **if** $Q_i^k \subseteq \tilde{Q}$ **then**
5:          Remove $\beta_i^k$ from the set $B_{-(x,y)}$;
6:       **else**
7:          $r(\beta_i^k) = \frac{c_i^k}{|Q_i^k - \tilde{Q}|}$;
8:       **end if**
9:    **end for**
10:    Sort $r(\beta_i^k)$ for all $\beta_i^k \in B_{-(x,y)}$ in the nondecreasing order and the list is denoted by $\mathcal{R}$;
11:    $\beta_s^t$ denotes the head of $\mathcal{R}$;
12:    **if** $\beta_s^t$ conflicts with existing winning bids **then**
13:       CONTINUE;
14:    **end if**
15:    **if** $Q_x^y \subseteq \tilde{Q} \bigcup Q_s^t$ **then**
16:       $c(\beta_x^y) \leftarrow \beta_s^t,\ p_c(\beta_x^y) \leftarrow r(\beta_s^t) \cdot |Q_x^y - \widetilde{Q}|$.
17:       RETURN $c(\beta_x^y),\ p_c(\beta_x^y)$;
18:    **end if**
19:    $\tilde{Q} \leftarrow \tilde{Q} \bigcup Q_s^t$;
20:    Remove $\beta_s^t$ from $B_{-(x,y)}$;
21: **end while**

**Definition 8** (**Critical Value**). *For each bid $\beta_i^k$, there is a critical value $\gamma_i^k$. If bid $\beta_i^k$ declares a cost that is lower than or equal to $\gamma_i^k$, it must win; otherwise; it will not win.*

Next, we prove *TRAC* is truthful by showing that it satisfies both the two conditions.

**Lemma 1.** *Algorithm 1 of TRAC is monotonic.*

*Proof:* Suppose bid $\beta_i^k$ wins in the $q$-th iteration. In the previous iterations, a number of winning bids have been determined. Let a sorted list $\mathrm{L} = (\xi^1, \xi^2, \cdots)$ storing these winning bids in the order that they have been determined. Suppose $\beta_i^k$ is in the $q$-th place in the list. Assume bid $\beta_i^k$ was replaced by another bid $\tilde{\beta}_i^k = (Q_i^k, \tilde{b}_i^k)$, where $\tilde{b}_i^k) = b_i^k - \delta$. According to the rule for determining a winning bid in Algorithm 1, bid $\tilde{\beta}_i^k$ must have won in the $q$-th or an even earlier iteration. This proves that Algorithm 1 is monotonic. ∎

**Lemma 2.** *Each winning bid is paid the critical value.*

*Proof:* If a winning bid $\beta_x^y$ is $(Q_x^y, p_c(\beta_x^y))$, the ranking criterion of Algorithm 2 in the $q$-th iteration is $r^q(\beta_x^y) = r^q(\beta_i^k)$. It is obvious that a bid $\tilde{\beta}_x^y = (Q_x^y, p_c(\beta_x^y) - \delta), \delta > 0$ whose claimed cost is lower than $\beta_x^y$ would win, because in the $q$-th iteration $r^q(\tilde{\beta}_x^y) < r^q(\beta_i^k)$. On the contrary, a bid $\tilde{\beta}_x^y = (Q_x^y, p_c(\beta_x^y) + \zeta), \zeta > 0$ whose claimed cost is higher than $\beta_x^y$ it would not win in the $q$-th iteration and could not win in the following iterations because $Q_x^y \subseteq \tilde{Q}^q$. This demonstrates that $p_c(\beta_x^y)$ is the critical value. ∎

**Theorem 2.** *TRAC is truthful.*

According to [13], this theorem easily follows from Lemma 1 and Lemma 2.

**Theorem 3.** *TRAC meets the condition of individual rationality.*

*Proof:* For a smartphone $x$ that has no winning bids, its payoff is zero, *i.e.*, $u_x = 0$. For a smartphone $x$ with winning bids, its payoff is $u_x = \sum_{\beta_x^y \in S} u(\beta_x^y)$. Next, we show that each $u(\beta_x^y)$ is nonnegative, $u(\beta_x^y) \geq 0$. Since bid $\beta_x^y$ is winning, its claimed cost must be smaller than its critical value, *i.e.*, $c_x^y \leq p_c(\beta_x^y)$. Then for each winning bid $\beta_x^y$ that truthfully report its real cost, its payoff is $u_x = p_c(\beta_x^y) - c_x^y = p_c(\beta_x^y) - c_x^y \geq 0$. ∎

### B. Computational Efficiency

We next analyze the computation complexity of the two algorithms of *TRAC*.

**Lemma 3.** *Algorithm 1 of TRAC for winning bids determination has polynomial-time computation complexity.*

*Proof:* The complexity of the first for loop (line 3-9) and the second for loop (line 14-18) in Algorithm 1 is $O(|B|)$. The operation of sorting in line 10 is $O(|B| \cdot log|B|)$. Thus, the aggregate complexity of a single iteration of the outer while is $O(|B| \cdot log|B|)$. Since the outer loop is run at most $|T|$ times, it is easy to compute the total computation complexity of Algorithm 1 which is $O(|T| \cdot |B| \cdot log|B|)$. ∎

**Lemma 4.** *Algorithm 2 of TRAC for critical payment determination has polynomial-time computation complexity.*

*Proof:* The outer while loop is run at most $|B|$ times because in some steps $\tilde{Q}$ does not expand. In each iteration of the outer while loop, there are two for loops and an operation of sorting. The computation complexity of the first for loop is $O(|B|)$. The computation complexity of the sorting operation is $O(|B|log|B|)$. The computation complexity of the second for loop (removing conflicting bids in line 12 ) is $O(|T|)$. Thus, the total computation complexity is $O(|B| \cdot (|B|log|B| + |T|))$. ∎

### C. Approximation Ratio Analysis

We finally analyze the approximation ratio achieved by Algorithm 1.

**Theorem 4.** *Algorithm 1 can approximate the optimal solution within a factor of $H(m)$, where $m$ is the maximum number of sensing tasks that a smartphone can accommodate,* i.e., *$m = max_{\beta_i^k \in B}\{|Q_i^k|\} \leq |T|$ and $H(m) = \sum_{i=1}^{m} \frac{1}{i} \approx \ln m$.*

*Proof:* When Algorithm 1 chooses a bid $\beta_i^k$, imagine that each element in the set $Q_i^k - \tilde{Q}$ introduces a part of the social cost. Then, the total social cost of all winning bids selected by Algorithm 1 equals the amount of the sum of social cost introduces in all iterations.

Consider an arbitrary winning bid $\beta_i^k = (Q_i^k, c_i^k)$, $Q_i^k = \{t_n, \cdots, t_1\}$, $\beta_i^k \in I$. Suppose that the elements of $Q_i^k$ is covered in the order of $t_n, t_{(n-1)}, \cdots$. At the beginning of iteration in which Algorithm covers $t_j$ of $Q_i^k$, at least $j$ elements of $Q_i^k$ is uncovered, *i.e.*, in $Q_i^k - \tilde{Q}$. Thus, if the algorithm chooses $\beta_i^k$ in that iteration, element $t_j$ introduces

TABLE I
SUMMARY OF DEFAULT SETTINGS

| Parameter name | Default value |
|---|---|
| Number of smartphones $n$ | 500 |
| Number of sensing tasks $m$ | 40 |
| Cost range $R$ | [0,50] |
| Maximum number $r_i$ of bids a user could receive | 3 |

at most $\frac{c_i^k}{j}$. Then, the social cost introduced by all elements in $Q_i^k$ is $\sum_{j=1}^{n} \frac{c_i^k}{j} = H(n)c_i^k$. Summing over each $\beta_i^k \in I$, the social cost obtained by Algorithm 1 is $\tilde{\omega} \leq \sum_{\beta_i^k \in I} H(n)c_i^k \leq H(m)\sum_{\beta_i^k \in I} c_i^k$ and $m = max\{|Q_i^k|\}$. The optimal social cost $\omega^*$ is $\sum_{\beta_i^k \in I} c_i^k$, and then, $\tilde{\omega} \leq H(m) \cdot \omega^*$. ■

## V. PERFORMANCE EVALUATION

### A. Methodology and Simulation Settings

We evaluate the performance of the proposed mechanisms with extensive simulations based on a real data set of location traces. The real location traces were collected from around 2,600 taxis in Shanghai, as used in prior studies [14] [15]. For each taxi, its GPS coordinate (longitude and latitude) and the corresponding ID were recorded every 30 to 60 seconds. The taxis operate in Shanghai, the largest city in China which covers an area of 635 $km^2$. In simulation, we take the locations of a subset of the taxis at a certain time snapshot. For different simulations, we take different snapshots. We assume that a smartphone is carried by the passenger or the driver of a selected taxi.

To evaluate the performance of *TRAC*, we use the following metrics: *social cost, overpayment ratio, individual rationality, running time and approximation ratio*. The overpayment is the difference between the total payment to all contributing smartphones and the social cost. The overpayment measures the cost paid by the platform (or smartphone sensing applications) to induce truthfulness of all smartphones. We define the overpayment ratio to measure the overpayment.

**Definition 9** (**Overpayment Ratio**). *The overpayment ratio is ratio of overpayment to the social cost. It is computed as*

$$\lambda = \frac{P - \omega}{\omega}, \tag{9}$$

*where $P$ denotes the total payment.*

In simulation, location attributes of sensing tasks are uniformly distributed in the whole area of the Shanghai. We generate real costs of bids according to three distributions, *i.e.*, uniform distribution (UNM), normal distribution (NORM) and exponential distribution (EXP). Each experiment is conducted with each of all three distributions. In simulation, we vary the mean $\mu$ of real costs from 15 to 25. The normal distribution sets such a standard deviation $\sigma$ that 80% of samples falling within $[\mu - \sigma, \mu + \sigma]$. The default settings of other parameters are summarized in Table I. Each data point is the average of 20 independent runs under the same setting.

### B. Evaluation of Overpayment Ratio

Fig. 4 plots the overpayment ratio when number of smartphones changes from 400 to 1,000. We can see that the overpayment ratio keeps low when the number of smartphones increases. The overpayment ratio is always lower than 1.5 for all three kinds of distributions. The overpayment ratio of the exponential distribution is larger than those of the other two distributions. This is because, with the exponential distribution, if the real cost of its critical bid locates in the tail of the exponential distribution, the overpayment is relatively large. The overpayment ratio of the normal distribution is slightly larger than that of the uniform distribution because with the normal distribution, the real cost of critical bid of the smartphone is closer to $\mu$ and is possibly larger than that of the uniform distribution.

Fig. 5 shows that with the increasing number of sensing tasks, the overpayment ratio fluctuates and then converges to 1.3 for all three distributions. The overpayment ratio remains small as the number of sensing tasks becomes larger.

Then, in Fig. 6, the overpayment ratio is evaluated when the average of real costs increases. The mean of real costs vary from 15 to 25. The overpayment ratio increases slightly with the average of real costs increases. From the figure, we can see that the overpayment does not exceed 2.

### C. Evaluation of Social Cost

Fig. 8 depicts the performance of social cost with the number of smartphones being varied from 400 to 1,000. The social cost decreases when the number of smartphones increases. This is because when there are more smartphones, the platform can find more cheap smartphones to perform the sensing tasks. The social cost of the exponential distribution is lower than those of the other two distributions since there exists a larger percent of smartphones of low real costs.

Fig. 9 plots the social cost when the total number of sensing tasks varies from 10 to 50. With more sensing tasks, the platform must employ more smartphones. Thus, more resources are consumed, incurring a higher social cost. The social cost of the exponential distribution is smaller than those of the other two distributions again because the exponential distribution produces more smartphones with low costs. When the number of sensing tasks $m$ is larger, the social cost of the uniform distribution is slightly larger than that of the normal distribution. When the platform has to choose a large number of smartphones, it is more likely to select cheaper ones when the real costs subject to normal distribution.

In Fig. 10, the social cost rises gently with the average real cost increases. When the average of real costs goes up, the collaborative sensing applications are less likely to find cheap smartphones, thus incurring more social cost. The increase of the exponential distribution is rather smaller since there are more cheaper smartphones than that of the other two distributions.

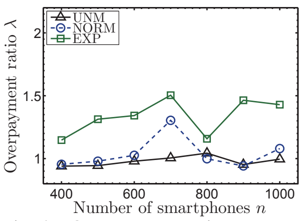

Fig. 4. Overpayment ratio vs. Number of smartphones $n$.

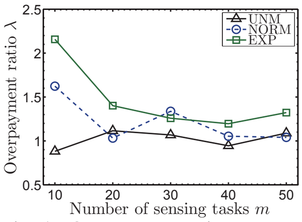

Fig. 5. Overpayment ratio vs. Number of sensing tasks $m$.

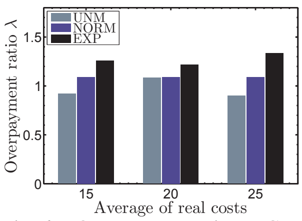

Fig. 6. Overpayment ratio vs. Cost range $R$.

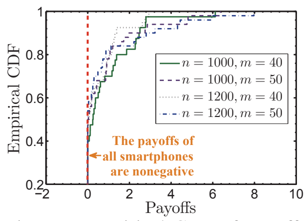

Fig. 7. Empirical CDF of payoffs for all smartphones.

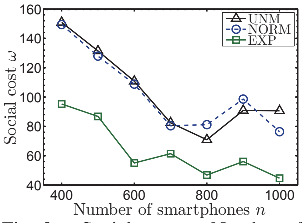

Fig. 8. Social cost vs. Number of smartphones $n$.

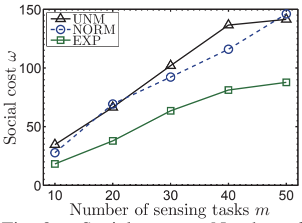

Fig. 9. Social cost vs. Number of sensing tasks $m$.

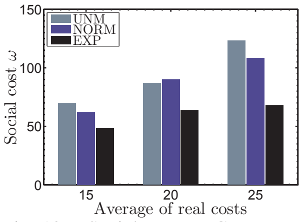

Fig. 10. Social cost vs. Cost range $R$.

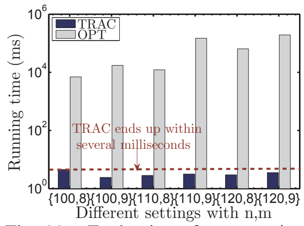

Fig. 11. Evaluation of computation efficiency (y-axis in log scale).

### D. Evaluation of Individual Rationality

The basic property of an incentive mechanism is to guarantee that each smartphone is individually rational. In Fig. 7, we plot the empirical CDF of the payoffs for all smartphones. From the figure, we can see that no smartphone has a negative payoffs, thus demonstrating that *TRAC* achieves the property of individual rationality.

### E. Evaluation of Computation Efficiency

We compare the running time of *TRAC* with the optimal VCG mechanism (denoted by *OPT*). Since the problem is *NP hard*, we can only obtain the optimal solution when the problem scale is small (*e.g.*, $n \leq 120, m \leq 9$). The *OPT* employs a backtracking approach to find the optimal solution and a VCG-style payment scheme to guarantee truthfulness. In Fig. 11, we show the running time in various settings. We can clearly see that our algorithm uses significantly shorter time than the optimal algorithm. For example, in the 4th set of bars, *TRAC* uses only 3 milliseconds, but *OPT* uses more than 2 minutes.

## VI. Related Work

We review related work from three aspects, including incentive schemes with strategic users, incentive schemes with cooperative users, and privacy preservation.

***Incentives mechanism with strategic behaviors***: In [2], two incentive mechanisms are designed for a user-centric model and a platform centric model, respectively. In the platform-centric model, the platform first announces the total amount of money that the platform is willing to pay, and then each smartphone decides the time it would like to provide the sensing service. In the user-centric model, the objective is system-wide. If a sensing task is finished, it would receive some benefit. The goal is to select a subset of smartphones, maximizing the overall gain. However, they only consider single preference of a user, which means a user only submits a single bid and it would win for performing all sensing tasks in the bid or not at all. Such an assumption is not realistic for location-aware collaborative sensing, failing to make the best use of the sensing capabilities of all smartphones. In [11], the author designs an optimal incentive mechanism for participatory sensing and aim to maximize the revenue of the service provider. The participating level of each user is characterized by units and the user has a cost for each unit of participation. The algorithm fails to take the location into consideration and could not be applied to our problem because of the NP-hardness of winning bids determination problem.

***Incentives schemes with cooperative users***: Some related work proposing to provide monetary rewards to generally cooperative users, such as [16] [17] [18]. Users are supposed to be well motivated because they receive some monetary rewards. In [16], a subset of users are greedily selected according to their locations and the total budget. The algorithm aims to cover the largest area. In [17], the users sell sensed data to a service provider and a dynamic pricing scheme is designed to stimulate more participants, thus achieving better quality of service. In [18], it considers that a selfish user has the demand of consuming data, and how much service it could consume depends on how much she or he contributes to the participatory sensing system. Thus, they have to consider how to satisfy all users fairly and how to achieve a desirable result (maximized social welfare) for the whole system. These existing studies do not consider that cost information is private, and that users may misreport their real costs in order to maximize their own payoffs. As a result, these incentive schemes are not truthful.

***Privacy preserving schemes***: Privacy preserving is crucial to mobile crowdsourcing with smartphones. It has attracted many research efforts, such as [9] [10] [19] [20]. One smartphone is supposed to report its sensed data, but it is reluctant to disclose its private information, such as location, and identity. In [21] and [22], the authors design regression models for data which allow accurate model construction while keeping individual information private. Some other studies

[23] [24] add noise or perturbation to original sensory data for the purpose of anonymity. These papers usually utilize k-anonymity or entropy to measure the leakage of privacy. Our work, which has focused on designing incentive mechanisms for location-aware collaborative sensing, can benefit from these existing schemes for protecting location privacy.

## VII. Conclusion and Future Work

In this paper, we have investigated the important problem of designing incentive mechanisms for stimulating smartphone users to join mobile crowdsourcing applications with smartphones. We have uniquely considered the *crucial dimension of location information* in the design of incentive mechanism. Based on the reverse auction framework, we have designed *TRAC* which consists of a near-optimal approximate algorithm and a critical payment scheme. With polynomial-time computation complexity, the algorithm for determining winning bids can approximate the optimal solution within a factor of $1+\ln(n)$, where $n$ is the maximum number of sensing tasks that a smartphone can accommodate. And although the algorithm of determining winning bids is approximate, the payment scheme can guarantee truthfulness. We have performed both rigid theoretical analysis and extensive simulations, and results have shown the proposed *TRAC* achieves the desired properties, such as truthfulness, individual rationality and high computation efficiency.

Based on the current work, we will carry out future work along the following directions. First, a smartphone may possibly misreport its own location in order to maximize its own payoff. We will extend our mechanism design to make misreporting into consideration. Second, since a smartphone should share its location information, this may introduce the problem of privacy breach. Privacy protection issues should be studied in mobile crowdsourcing. Finally, several smartphones may collude to gain better payoffs. We will propose to extend *TRAC* to resist collusion attacks.

## Acknowledgements

This research is supported by China 973 Program (2014CB340303 and 2013CB329006), NSFC (No. 61170238, 60903190, 60933011, 61173156, 61202375), Doctoral Fund of Ministry of Education of China (20100073120021), National 863 Program (2009AA012201, 2011AA010500 and 2013AA01A601), SJTU SMC Project (201120), STCSM (08dz1501600, 12ZR1414900), Singapore NRF (CREATE E2S2), RGC (CERG 622410, 622613 and HKUST6/CRF/12R) and MSRA funding for Urban Computing and for Star Track program. This work is also supported by Program for Changjiang Scholars and Innovative Research Team in University (IRT1158, PCSIRT), China.

## References

[1] “Sensorly”, http://www.sensorly.com/.

[2] Dejun Yang, Guoliang Xue, Xi Fang, and Jian Tang, “Crowdsourcing to smartphones:incentive mechanism design for mobile phone sensing”, in *ACM MOBICOM*, 2012.

[3] Mohammad Nozari Zarmehri and Ana Aguiar, “Supporting sensing application in vehicular networks”, in *ACM CHANTS*, 2012.

[4] Raghu K Ganti, Fan Ye, and Hui Lei, “Mobile crowdsensing: Current state and future challenges”, *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[5] Min Mun, Sasank Reddy, Katie Shilton, Nathan Yau, Jeff Burke, Deborah Estrin, Mark Hansen, Eric Howard, Ruth West, and Pter Boda, “Peir: the personal environmental impact report, as a platform for participatory sensing systems research”, in *ACM MobiSys*, 2009.

[6] Rajib Kumar Rana, Chun Tung Chou, Salil S. Kanhere, Nirupama Bulusu, and Wen Hu, “Ear-phone: an end-to-end participatory urban noise mapping system”, in *ACM/IEEE IPSN*, 2010.

[7] Arvind Thiagarajan, Lenin Ravindranath, Katrina LaCurts, Samuel Madden, Hari Balakrishnan, Sivan Toledo, and Jakob Eriksson, “Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones”, in *ACM SenSys*, 2009.

[8] Anshul Rai, Krishna Kant Chintalapudi, Venkata N.Padmanabhan, and Rijurekha Sen, “Zee: Zero-effort crowdingsourcing for indoor localization”, in *ACM SIGCOMM*, 2012.

[9] Xielei(Oscar) Wang, Wei Cheng, Prasant Mohapatra, and Tarek Abdelzaher, “Artsense: Anonymous reputation and trust in participatory sensing”, in *IEEE INFOCOM*, 2013.

[10] Alfred C.Weaver, Joseph P.Boyle, and Liliya I.Besaleva, “Application and trust issues when crowdsourcing a crisis”, in *IEEE ICCCN*, 2012.

[11] Iordanis Koutsopoulos, “Optimal incentive-driven design of participatory sensing systems”, in *IEEE INFOCOM*, 2013.

[12] Yu Zhang and Mihacla van der Schaar, “Reputation-based incentive protocols in crowdsourcing applications”, in *IEEE INFOCOM*, 2012.

[13] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani, *Algorithmic game theory*, Cambridge University Press, 2007.

[14] Yuchen Wu, Yanmin Zhu, and Bo Li, “Infrastructure-assisted routing in vehicular networks”, in *INFOCOM*, 2012.

[15] Yuchen Wu, Yanmin Zhu, Hongzi Zhu, and Bo Li, “Ccr: Capacity-constrained replication for data delivery in vehicular networks”, in *INFOCOM*, 2013.

[16] Luis G.Jaimes, Idalides Vergara-Laurens, and Miguel A.Labrador, “A location-based incentive mechanism for participatory sensing systems with budget constraints”, in *IEEE percom*, 2012.

[17] Juong-Sik Lee and Baik Hoh, “Sell your experiences:a market mechanism based incentive for participatory sensing”, in *IEEE percom*, 2009.

[18] Afra J. Mashhadi, Sonia Ben Mokhtar, and Licia Capra, “Fair content dissemination in participatory dtns”, *Ad Hoc Networks*, vol. 10, no. 8, pp. 1633–1645, 2012.

[19] Michael M.Groat, Benjamin Edwards, James Horey, Wenbo He, and Stephanie Forrest, “Enhancing privacy in participatory sensing applications with multidimensional data”, in *IEEE percom*, 2012.

[20] Harald Weinschrott, Julian Weisser, Frank Durr, and Kurt Rothermel, “Participatory sensing algorithms for mobile object discovery in urban areas”, in *IEEE percom*, 2011.

[21] Hossein Ahmadi, Nam Pham, Raghu Ganti, Tarek Abdelzaher, Suman Nath, and Jiawei Han, “Privacy-aware regression modeling of participatory sensing data”, in *ACM Sensys*, 2010.

[22] Kai Xing, Zhiguo Wan, Pengfei Hu, Haojin Zhu, Yuepeng Wang, Xi Chen, Yang Wang, and Liusheng Huang, “Mutual privacy-preserving regression modeling in participatory sensing”, in *IEEE INFOCOM*, 2013.

[23] Fan Zhang, Li He, Wenbo He, and Xue Liu, “Data perturbation with state-dependent noise for participatory sensing”, in *IEEE INFOCOM*, 2012.

[24] Khuong Vu, Rong Zheng, and Jie Gao, “Efficient algorithms for k-anonymous location privacy in participatory sensing”, in *IEEE INFOCOM*, 2012.