# Quality-Driven Auction-Based Incentive Mechanism for Mobile Crowd Sensing

Yutian Wen, Jinyu Shi, *Student Member, IEEE*, Qi Zhang, Xiaohua Tian, *Member, IEEE*, Zhengyong Huang,
Hui Yu, Yu Cheng, *Senior Member, IEEE*, and Xuemin (Sherman) Shen, *Fellow, IEEE*

*Abstract*—The recent paradigm of mobile crowd sensing (MCS) enables a broad range of mobile applications. A critical challenge for the paradigm is to incentivize phone users to be workers providing sensing services. While some theoretical incentive mechanisms for general-purpose crowdsourcing have been proposed, it is still an open issue as to how to incorporate the theoretical framework into the practical MCS system. In this paper, we propose an incentive mechanism based on a quality-driven auction (QDA). The mechanism is specifically for the MCS system, where the worker is paid off based on the quality of sensed data instead of working time, as adopted in the literature. We theoretically prove that the mechanism is truthful, individual rational, platform profitable, and social-welfare optimal. Moreover, we incorporate our incentive mechanism into a Wi-Fi fingerprint-based indoor localization system to incentivize the MCS-based fingerprint collection. We present a probabilistic model to evaluate the reliability of the submitted data, which resolves the issue that the ground truth for the data reliability is unavailable. We realize and deploy an indoor localization system to evaluate our proposed incentive mechanism and present extensive experimental results.

*Index Terms*—Incentive, indoor localization, mobile crowd sensing (MCS).

## I. INTRODUCTION

**M**OBILE phones have become increasingly more intelligent over the past few years, which not only own the processing power comparable to that of laptops but accommodate a rich set of sensors such as an accelerometer, a compass, a gyro-

Y. Wen, J. Shi, Q. Zhang Z. Huang, and H. Yu are with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: nogerw@sjtu.edu.cn; mizaoyu@sjtu.edu.cn; 362802781@sjtu.edu.cn; zhengyonghuang@sjtu.edu.cn; yuhui@sjtu.edu.cn).

X. Tian is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China, and also with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: xtian@sjtu.edu.cn).

Y. Cheng is with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL 60616 USA (e-mail: cheng@iit.edu).

X. Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: xshen@bbcr.uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TVT.2014.2363842

scope, a GPS, a microphone, and a camera as well. With appropriate organization, mobile phones could form sensing networks enabling new mobile applications across various domains [1]. For example, GPSs in mobile phones can be utilized to collect traffic information and help users estimate travel time [2], phone sensors can help track the individual behavior to evaluate the impact on the environment pollution [3], and phone-embedded microphones can help create noise maps in different areas [4].

The sensing paradigm of utilizing the crowd power to fulfill large-scale sensing tasks at a low cost is termed as *mobile crowd sensing* (MCS). A MCS application of particular interest is that the sensors embedded in mobile phones are leveraged for various sensing tasks [14]. Compared with the traditional paradigm relying on wireless sensor networks, MCS has broader coverage, lower deployment cost, and higher scalability. The MCS system basically consists of the mobile phone users acting as sensing service providers (workers), the data requesters who want to get data from users, and an agent platform acting as a medium to recruit workers for requesters to perform data collecting tasks. Most of the existing MCS systems recruit volunteers as workers [10], although performing the sensing task will consume workers' phone resources and potentially incur privacy leakage. However, fully exploiting the potential of distributed mobile phone resources needs a large amount of participants. Consequently, designing a proper incentive mechanism for the MCS system is vitally important.

Game theory is used to address the issue because of its straightforward suitability for modeling the trading process [5]– [7]. The Stackelberg game, contract theory, and auction theory are employed to model the interactions between workers and the platform [8]–[10], [12]. While these models can be theoretically proved having favored characteristics such as truthfulness and profitability, putting the theory into practice is hardly straightforward. In most of the work in the literature, workers are paid off by the workload undertaken, which is usually evaluated by the working time. Nevertheless, the working-time-based evaluation is unable to fit all MCS scenarios, particularly for some data collection system where the quality of the submitted data instead of the working time is more important. However, evaluating the quality of the crowd-sensed data itself is nontrivial [13] because there is usually no perfect benchmark to measure the reliability of the crowd-sensed data. The challenge for the practical incentive mechanism design for the MCS system is twofold: 1) The theoretical framework to model the actual interaction between workers and the platform is still incomplete, and 2) the effective approach to evaluate the quality of the crowd-sensed data needs more investigation.

This paper studies how to design the incentive mechanism for data collecting MCS systems, with the indoor localization system as an example. The indoor localization becomes increasingly popular with the rise of location-based services [20]–[22], where collecting the received signal strength (RSS) values of those Wi-Fi access points (APs) within the building is an important part [23], [24]. Since collecting such a large amount of data could be expensive and laborious for any single entity, collecting RSS with the methodology of MCS has been acknowledged [15], [17], [18]. However, the state of the art of incorporating the MCS into the RSS collecting process is still at the concept level; most of the existing studies still focus on the localization technique itself. The fundamentals and the details of the incentive mechanism design for MCS are still not clear.

In this paper, we propose an incentive mechanism to model the interaction between workers and the platform, with the quality of submitted data taken into account. A probabilistic model is presented to evaluate the reliability of the crowd-sensed data in the indoor localization scenario. Specifically, our contributions are as follows.

- We design an incentive mechanism for the MCS system based on a quality-driven auction (QDA). The requesters post the task over the agent platform, and the interested worker submits the collected data and corresponding price. We present an effective algorithm to select a group of data that can maximize the social welfare of the system. We show that the proposed mechanism can encourage more submissions of high-quality data with lower computational complexity. In our mechanism, the platform has no need to know the cost of each individual worker, which is supposed to be private information. We theoretically prove that the proposed mechanism is truthful, individual rational, platform profitable, and social-welfare optimal.
- We propose a probabilistic model to evaluate the reliability of the MCS data in the scenario of indoor localization. We transform the unreliability of the data into the unreliability of the human's sense of locality, which can be profiled by prior experiments of the human behavior once and for all. The profile gives the probability of a human's incorrectness of locality, which is then utilized to find the submitted data with the highest reliability. The reliability evaluation scheme is smoothly integrated into the proposed incentive mechanism, with all the benefits reserved.
- We develop and deploy a practical indoor localization system covering over 100 m$^2$ of a building at the Shanghai Jiao Tong University campus. The system is constructed following the mobile cloud architecture. The mobile users collect the RSS information and transmit the data to the cloud, which is implemented with CloudFoundry [25]. Extensive experimental results are presented to illustrate the performance of our scheme.

## II. RELATED WORK

### A. General-Purpose Incentive Mechanisms

Models in game theory can be borrowed to design the incentive mechanism. Yang *et al.* proposed two types of incentive mechanisms for the MCS system in the perspectives of the

agent platform and mobile users, respectively [8]. The platform-centric mechanism is based on the Stackelberg game, where it is assumed that the agent platform has the absolute control over the total payment to users who can only adjust their strategies to comply. The user-centric incentive mechanism utilizes an auction-based scheme and owns benefits such as truthfulness.

Duan *et al.* classified the MCS system into two classes: data acquisition and distributed computing [9]. The former serves the purpose of collecting data for building up a database, and the latter utilizes distributed computation power to solve problem that could be expensive for a single device. The Stackelberg game is used to model the interaction between workers and requesters in the data acquisition scenario, and the contract theory is applied in the distributed computing scenario where the complete information and incomplete information settings are considered.

The Stackelberg game model needs the platform to know the information of users in advance, which is too strong in the practical system. The auction-based model in the literature, however, has not taken the data quality into consideration.

### B. Incentive Mechanisms for Specific Purposes

Zhao *et al.* proposed an online incentive mechanism for the scenario where workers arrive one by one [10], which is in contrast to some mechanisms assuming all of workers report their profiles to the agent platform in advance. The problem is modeled as an online auction, where mobile users submit their private information to the platform over time and a subset of users are selected before a specified deadline.

To shorten the crowd response time, Bernstein *et al.* proposed the retainer model, where workers are recruited in advance and held idle for a small amount of expense called retainer. The reserved workers will respond quickly when tasks are assigned [11]. Based on the retainer model, Patrick *et al.* proposed a combinatorial allocation and pricing scheme for crowdsourcing tasks with time constraints [12]. The workers are selected from all possible candidates with an optimization-based procedure, and the payments for workers are calculated using a Vickrey–Clarke–Groves-based rule.

Although referring to the reverse Vickrey auction model, our scheme considers the reliability of the submitted data, which provides higher efficiency of funding utilization. The experimental results will show that our scheme can select more proportion of reliable data with limited computation time.

### C. Incentive Mechanisms of MCS for Indoor Localization

Many research efforts have been dedicated to the indoor localization over the past decades, among which the Wi-Fi fingerprint-based methodology inspired much work due to its outstanding balance between accuracy and simplicity [16]–[18], [20]. The fingerprint-based localization technique can be divided into the training and localizing phases. In the training phase, the fingerprints at positions of interest are collected into a database by measuring the RSS of Wi-Fi APs around. In the localizing phase, the system will search the user's current fingerprint in the fingerprints database and return the optimally

matched location. Collecting such a large amount of data in the training phase could be extremely expensive and laborious by well-trained experts; therefore, offloading the RSS collection to the MCS becomes a consensus [15], [18], [20] but has barely been investigated.

The challenge of designing an incentive mechanism for MCS RSS is that the collected data could be noisy and unreliable since workers are usually paid with very limited reward and the tasks performed are usually monotonous. It is not always appropriate to evaluate the worker's contribution simply by the working time as in most of the existing work mentioned above; however, evaluating the reliability of submitted data itself is nontrivial because there is no perfect benchmark. Karger *et al.* studied the reliability issue for the question-and-answer crowdsourcing system, which shows that it is possible to obtain a correct answer to each task with certain probability [13]. The correctness probability is guaranteed at the cost that each task has to be assigned to multiple workers. However, it is extremely difficult to apply the theoretical results to the practical MCS system because guaranteeing enough redundancy for each task and certain characteristics of users as required by the model could be hard in practice. He *et al.* proposed a pricing mechanism based on bargaining theory for MCS, which realistically considers the task performing cost and the market demand [14]. This paper however focuses on the influence of data quality on the payoffs of the data collecting workers.

## III. SYSTEM MODEL AND DESIGN CHALLENGES

### A. System Model

We consider the MCS system consisting of three kinds of players: workers, agent platform, and requesters. The platform aggregates the demands from different requesters, recruits workers, checks the reliability of submitted data, and supplies the selected data to requesters. The time is slotted in the model, and the following processes will be performed for each time slot:

- contract determination;
- winner data set determination;
- payment determination;
- response and update.

The main notations used in this paper are tabulated as in Table I.

*Contract determination*: After receiving the demand, a set of workers $N = \{1, 2, \ldots, n\}$ will collect and submit the requested data with each claiming a price $b_{ij}$ for every kind of submitted data $x_{ij}$, where $x_{ij}$ represents the data of type $j$ collected by user $i$ and $b_{ij}$ is the lowest acceptable payment user $i$ asks for submitting $x_{ij}$. A worker could collect as many kinds of information as possible. We use $M = \{1, 2, \ldots, m\}$ to denote all types of data needed to be sensed in the system, where $M_i$ is a subset of $M$ containing the data types measured by user $i$. Collecting the data of each type is considered one task, and there are a total of $m$ tasks here. Each pair of data and claimed price is termed as a *subcontract* denoted as $c_{ij}$, and all subcontracts of worker $i$ is termed as a *contract*, i.e., $c_{ij} = (b_{ij}, x_{ij})$, $C_i = \{c_{ij} | j \in M_i\}$. The worker has no need to know other workers' claimed prices and just needs to wait for the response from the platform after uploading the contract.

TABLE I
MAIN NOTATIONS

| | |
|---|---|
| $n$ | Number of interested workers |
| $N$ | Set of interested workers |
| $m$ | Number of data types |
| $M$ | Compete set of data types |
| $M_i$ | Subset of $M$, containing all data types sensed by worker $i$ |
| $F$ | Set of all submitted data. |
| $x_{ij}$ | Submitted data of type $j$ measured by worker $i$ |
| $b_{ij}$ | Worker $i$'s lowest acceptable payment for $x_{ij}$ |
| $C_i$ | Contract set offered by user $i$ |
| $c_{ij}$ | Sub-contract offered by worker $i$ for data $x_{ij}$ |
| $c_{-ij}$ | All sub-contracts except the one for data $x_{ij}$ |
| $N^*$ | Winner user set, $N^* \subseteq N$ |
| $M_i^*$ | Types of winner data set measured by worker $i$ |
| $W$ | Winner data set, $W = \{x_{ij} | i \in N^*, j \in M_i^*\}$ |
| $W^{st}$ | Winner data set when data $x_{st}$ is not in the winner set |
| $k_{ij}$ | Cost for $x_{ij}$ |
| $p_{ij}$ | Payment to user $i$ for data $x_{ij}$ |
| $P_i$ | Set of $p_{ij}$, where $j \in M_i$ |
| $u_i$ | Utility of worker $i$ |
| $u_{ij}$ | Utility of worker $i$ for data $x_{ij}$ |
| $u_p$ | Utility of the platform |
| $L(x_{ij})$ | The value of data $x_{ij}$ |
| $R(\cdot)$ | Revenue function |
| $f(W)$ | Social welfare function |
| $f_p(W)$ | Social welfare function in platform's perspective |

*Winner data set determination*: The platform needs to determine a winner data set $W_j$ for each data type $j$ after receiving all submitted contracts. We use $F_j = \{x_{ij} | i \in N\}$ to represent the set of all submitted data of type $j$. The winner data set is the set that can result in the maximum social welfare denoted as $W_j = \text{argmax}\{f(W_j) | W_j \in F_j\}$, where $f(W_j)$ denotes the system social welfare of all data in $W_j$. We use $N_j^*$ to denote the set of the winners who have data of type $j$ being accepted by the requester. Thus, the winner data set is $W_j = \{x_{ij} | i \in N_j^*\}$, and we define $M_i^*$ as the set of data types that the worker $i$ collected and accepted by the requester.

*Payment determination*: After determining the winner data set, the platform needs to calculate the payment $p_{ij}$ that the requester should pay for each accepted data $x_{ij}$. If $x_{ij}$ is not in the winner data set, then $p_{ij} = 0$; otherwise, $p_{ij} > 0$ and $p_{ij}$ should be no less than $b_{ij}$ denoting the claimed price. Note that this is only the payment set given by one requester, and $p_{ij}$ may be different for different requesters. To incentivize workers to submit high-quality data, we use $k_{ij}$ to denote the cost of user $i$ if submitted data $x_{ij}$ are accepted by a single requester, and use $l \cdot k_{ij}$ to denote the cost if $x_{ij}$ are accepted by $l$ requesters. This is to offer the high-quality data provider high reward.

*Response and update*: Finally, the platform should pay off all workers for all data that they have submitted and data that are accepted by requesters. The platform will respond to user $i$ with a payment set $P_i = \{p_{ij} | j \in M_i\}$. After that, the accepted data will be adopted into the databases of corresponding requesters. The rejected data could be used for the next-round auction and may get accepted by requesters with different requirements on data quality.

Since workers are paid off according to the quality of submitted data, the proposed model can encourage participants of higher quality who are able to submit higher quality data. With workers submitting low-quality data get lower or even no

reward, the high-quality workers can get higher reward; thus, the utilization of the rewarding resource can be more effective.

### B. Design Challenges

Since workers are individual entities, it is difficult to ask them to negotiate with each other in practice. Moreover, the platform is normally unable to know the private information of workers in advance. The distributed scenario seems to be suitable for the auction model, specifically, the reverse Vickrey auction, the essence of which is simple: First, the platform searches all $2^n$ possible candidate winner data sets and calculates the total value and the social welfare that each candidate set will bring. Second, compare the total value brought by each candidate data set with the platform's required budget; if the total value of the data set is less than the budget, abandon this set. Third, sort all the remaining candidates by the social welfare that each can bring, and choose the candidate that can bring the greatest social welfare as the final winner data set.

However, the reverse Vickrey auction model has the following drawbacks if it were applied to the MCS system, which may hinder itself from being adopted.

1) All the data (replaceable items) will be regarded as the same in the reverse Vickrey auction; however, the crowd-sensed data for a single task in fact vary in their qualities. An effective incentive mechanism is supposed to encourage adoption of high-quality data.
2) The reverse Vickrey auction model will assume the cost of the worker as the quality of data submitted by the worker, which is not always the case in the MCS system. The worker may take many resources to collect some data, but the submitted data can turn out to be of low quality.
3) The platform will have to buy a certain amount of data even if the quality of the data is poor, which incurs inefficiency of funding utilization. This is because the platform could have saved the funding for higher quality data, instead of buying a group of low-quality data with a low value only to consume up the budget.
4) The social welfare in the model will only consider the workers' utilities and the platform's payment, which in together finally equals to the total cost of data in the winner set. It does not take the platform's revenue into account.
5) The model normally will have high computational complexity. This is because we have to search all possible combinations of submitted data to find the winner data set. The computation complexity is basically $O(2^n)$ if there are $n$ submitted data.

The fundamental reason of these drawbacks is that the reverse Vickrey model has to maintain the truthfulness property, which means that workers' claimed prices are their true costs for sensing the data, but this will lead to the fact that the payment of the platform must be independent of the prices asked by workers. To guarantee the independence, the platform has a fixed required budget that must be spent. The required budget here is the lowest value the platform should obtain from those selected data, which is oblivious to the actual quality of submitted data by workers.

## IV. QUALITY-DRIVEN AUCTION

### A. Overview

The idea of *QDA* is as follows. First, calculate a particular value for each subcontract, which reflects the extent to which the data are worth of buying and sort all subcontracts by that value. Second, separate the data into three categories and narrow down the searching range so that the candidate winner data are only selected from that range. Third, choose the data set that can maximize the social welfare of the system from the chosen range. The significant difference between the QDA and the reverse Vickrey auction is that we consider the revenue of the platform when calculating the social welfare and we do not need to have a required budget that must be spent, which can avoid buying low quality data. As we narrow down the searching range, the time spent on the winner data set determination will also sharply decrease. Moreover, QDA has the following favored properties.

*Individual rationality*: The worker whose submitted data are accepted by requesters will have a utility greater than 0. That is, any worker $i$'s utility for performing tasks is

$$u_i = \sum_{j \in M_i^*} u_{ij} = \sum_{j \in M_i^*} p_{ij} - \sum_{j \in M_i^*} k_{ij} \geq 0. \qquad (1)$$

*Truthfulness*: No worker can achieve a better utility by submitting the lowest acceptable payment other than its cost. Specifically, for any $i \in N$, $j \in M_i$ and any $b_{ij}$ other than $k_{ij}$, we have

$$u_{ij}(c_{ij}, c_{-ij}) = u_{ij}\left((b_{ij}, x_{ij}), c_{-ij}\right) \\ \leq u_{ij}\left((k_{ij}, x_{ij}), c_{-ij}\right) \qquad (2)$$

where $c_{ij}$ is user $i$'s strategy for data $x_{ij}$, and $c_{-ij}$ is the strategy profile excluding user $i$'s strategy for data $x_{ij}$.

*Platform profitability*: The utility of the platform $u_p$ is greater than zero, i.e.,

$$u_p = R\left(\sum_W L(x_{ij})\right) - \sum_W p_{ij} \qquad (3)$$

where $L(x_{ij})$ is an evaluation to the quality of data $x_{ij}$. We may consider $L(x_{ij})$ as the value of the data to a requester, and $R(\cdot)$ is the revenue function with the following properties: $R(0) = 0$, $R'(x) > 0$, $R''(x) < 0$. This is because adding reliable data $(L(x_{ij}) > 0)$ into the winner data set will always bring the platform benefit. With increasingly more reliable data accepted, the marginal revenue brought by new data will decrease. Consequently, the platform has decreasing marginal revenue.

*Social welfare maximization*: The total payoffs across all players are maximized. This means that all players including both workers and the platform are taken into account, in contrary to most of the work in the literature, which only focuses on either of them. We use the social welfare function $f(W)$ to quantify the social welfare as follows:

$$f(W) = \sum_{i \in N} u_i + u_p$$

$$= \sum_W (p_{ij} - k_{ij}) + R\left(\sum_W L(x_{ij})\right) - \sum_W p_{ij}$$

$$= R\left(\sum_W L(x_{ij})\right) - \sum_W k_{ij}. \qquad (4)$$
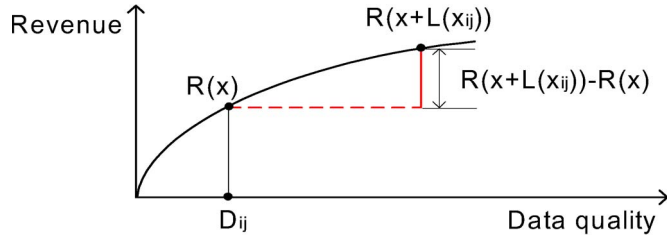
Fig. 1. $D_{ij}$.

We use $W^{st}$ to represent the winner data set where data $x_{st}$ are definitely rejected, i.e.,

$$f(W^{st}) = R\left(\sum_{W^{st}} L(x_{ij})\right) - \sum_{W^{st}} k_{ij}. \tag{5}$$

Note that the actual cost is only known by the worker himself; thus, the platform simply treats the lowest acceptable payment $b_{ij}$ as the cost for sensing data $x_{ij}$. Consequently, the social welfare in the platform's perspective is

$$f_p(W) = R\left(\sum_{W} L(x_{ij})\right) - \sum_{W} b_{ij}. \tag{6}$$

If data $x_{st} \in W$, then its payment will be

$$p_{st} = f_p(W) - f_p(W^{st}) + b_{st} \tag{7}$$

meaning that the incremental contribution data $x_{st}$ affects the whole system. However, if data are not accepted, then its payment will be zero.

It is worth mentioning that there may exist more than one winner sets, i.e., $\exists W_1$, $\exists W_2$, and $W_1 \neq W_2$, for any other $W$, $f(W_1) = f(W_2) \geq f(W)$. All these data sets are acceptable to the platform, and none of them violates the rule of payment. It is easy to prove that choosing any one of those winner sets will not hinder the truthfulness and individual rationality of QDA. If the platform chooses $W_2$ instead of $W_1$, apparently it will not affect those who are selected in both and those selected in neither. If $x_{ij} \in W_2$, $x_{ij} \notin W_1$, $p_{ij} = f(W_1) - f(W^{ij}) + b_{ij}$. Now that $x_{ij} \in W_2$, $x_{ij} \notin W_1$, $f(W_1) = f(W^{ij})$, $p_{ij} = b_{ij}$. This means that all users will only claim $b_{ij} = k_{ij}$, and the utility for the data $x_{ij}$ is always zero, regardless of whether the data are selected or not.

### B. Particular Value of the Sub-Contract

The first step of the QDA is to calculate the particular value of each subcontract mentioned earlier and sort all subcontracts by the value. This value is a measurement of to what extent the data are worth buying, which is influenced by both the data quality and the price. We use $D_{ij}$ to denote the value. Formally, if $R(L(x_{ij})) \geq b_{ij}$, then

$$D_{ij} = \max\{x | R(x + L(x_{ij})) - R(x) - b_{ij} \geq 0\}.$$

We here explain the meaning of $D_{ij}$ with Fig. 1. The horizontal axis means the data quality, and the vertical axis means that the revenue can be obtained by the platform. The curve stands for $R(\cdot)$, which is the revenue function. For data $x_{ij}$, we use the length of a line segment to represent its associated quality $L(x_{ij})$, such as the length of the horizontal dashed

line segment in Fig. 1. The starting and ending points of the dashed line segment are associated with two values $R(x)$ and $R(x + L(x_{ij}))$ on the curve. The increment of revenue by buying the data $x_{ij}$ can be measured as $R(x + L(x_{ij})) - R(x)$. The utility of buying the data $x_{ij}$ is $R(x + L(x_{ij})) - R(x) - b_{ij}$, where $b_{ij}$ is the price of the data. For a given $x_{ij}$, the horizontal distance of the two points on the curve is fixed, as well as $b_{ij}$. If we move the two points from left to the right on the curve while keeping their relative horizontal distance, the value $R(x + L(x_{ij})) - R(x) - b_{ij}$ is decreasing and finally will drop below zero because $R''(x) < 0$. The value $D_{ij}$ is the largest $x$ that can make the condition $R(x + L(x_{ij})) - R(x) - b_{ij} \geq 0$ still hold. We can see that each data will have an associated $D_{ij}$, which is only dependent on $R(\cdot)$, $L(x_{ij})$, and $b_{ij}$ and is independent of other submitted data.

If $R(L(x_{ij})) < b_{ij}$, $D_{ij} = 0$, which means that the revenue data $x_{ij}$ can bring to the platform is even lower than its own cost, adding data $x_{ij}$ to any set will make the social welfare decrease.

According to the definition above, if data $x_{ij}$ has a larger $D_{ij}$, the total value of the data that can be put into the winner set before $x_{ij}$ is selected is larger. Since the revenue function $R(\cdot)$ is monotonically increasing and adding $x_{ij}$ will not attenuate the social welfare, the platform can achieve higher social welfare. Consequently, data with larger $D_{ij}$ are worthier to be bought.

With the definition of $D_{ij}$, we can find many attributes of subcontracts, which can be used in the following description.

*Lemma 1:* For all $H \subset F, x_{ij} \in H$, if $\sum_H L(x_{ij}) > D_{ij} + L(x_{ij})$, then $f(H/x_{ij}) > f(H)$.

*Proof:* According to the definition, $D_{ij} + L(x_{ij})$ is already the largest value after worker $i$'s contribution, and it will not decrease the social welfare; however, if there is a set $H$ that has larger welfare than the former one, this means data $x_{ij}$ actually make the social welfare lower. ∎

*Lemma 2:* For all $H \subset F, x_{ij} \notin H$, if $\sum_H (L(x_{ij})) < D_{ij}$, then $f(H \cup x_{ij}) > f(H)$.

*Proof:* Since $R'(\cdot)$ is monotonically decreasing, adding data $x_{ij}$ to a set with a smaller total value will have a higher marginal revenue, although the cost remains the same, which will lead to higher social welfare. Consequently, adding $x_{ij}$ into a set whose total value is $D_{ij}$ will not decrease the social welfare, and adding it to a set with a smaller total value will have even larger social welfare. ∎

We assume that $W$ is a winner data set, and let $L = \sum_W L(x_{ij})$ be the total value of data in the winner data set. $\forall x_{ij} \in F \ \forall G \in R$, we divide the data set $F$ into three sets:

$$Q_1(G) = \{x_{ij} | D_{ij} + L(x_{ij}) < G\}$$
$$Q_2(G) = \{x_{ij} | D_{ij} < G < D_{ij} + L(x_{ij})\}$$
$$Q_3(G) = \{x_{ij} | G < D_{ij}\}.$$

Note that $Q_1 \cap Q_2 = Q_2 \cap Q_3 = Q_3 \cap Q_1 = \emptyset$, and $Q_1 \cup Q_2 \cup Q_3 = F$.

*Theorem 1:* $Q_1(L) \cap W = \emptyset$, and $Q_3(L) \subseteq W$.

*Proof:* According to Lemma 1, if $D_{ij} + L(x_{ij}) < L$, then $x_{ij} \notin W$ and $Q_1(L) \cap W = \emptyset$; according to Lemma 2, if $D_{ij} > L$, then $x_{ij} \in W$ and $Q_3(L) \subseteq W$. ∎
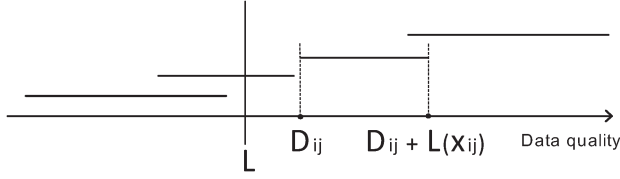
Fig. 2.　Algorithm 1.

*Theorem 2:*

- If $G > \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij})$, then $L < G$.
- If $G < \sum_{Q_3(G)} L(x_{ij})$, then $L > G$.
- If $\sum_{Q_3(G)} L(x_{ij}) < G < \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij})$, then $\min\{D_{ij}|x_{ij} \in Q_2(G)\} < L < \max\{D_{ij} + L(x_{ij})|x_{ij} \in Q_2(G)\}$.

*Proof:*

- If $G > \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij})$ and $L \geq G$, then $L > \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij})$. $W \subseteq Q_2(L) \cup Q_3(L)$ because $W$ is the winner set. $Q_2(L) \cup Q_3(L) \subseteq Q_2(G) \cup Q_3(G)$ because $L > G$. $L = \sum_W L(x_{ij}) \leq \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij}) < G$, which is contradictory to $L \geq G$.
- Similar to the proof above.
- We assume that $\sum_{Q_3(G)} L(x_{ij}) < G < \sum_{Q_2(G) \cup Q_3(G)} L(x_{ij})$. If $L$ is larger than $G$, and $W \cap Q_2(G) = \emptyset$, apparently $W \cap Q_1(G) = \emptyset$; therefore, $W \subseteq Q_3(G)$. Consequently, $\sum_{Q_3(G)} L(x_{ij}) \geq L > G$, which is a contradiction. Therefore, if $L$ is larger than $G$, then $W \cap Q_2 \neq \emptyset$. To keep at least one element of $Q_2(G)$ in $W$, there must exist at least one element with $D_{ij} \in Q_2(G)$ with $L < D_{ij} + L(x_{ij})$; thus, $L < \max\{D_{ij} + L(x_{ij})|x_{ij} \in Q_2(G)\}$. The proof for the case when $L$ is smaller than $G$ is likewise. If $L = G$, then $\sum_{Q_3(L)} L(x_{ij}) < L < \sum_{Q_2(L) \cup Q_3(L)} L(x_{ij})$; therefore, $W \cap Q_2(G) \neq \emptyset$. ∎

### C. Algorithm of QDA

We present the algorithm of determining the winner data set for a specific type of task for the convenience of presentation. The process of determining the entire submitted data set is similar and thus omitted here.

*Theorem 3:* The output of Algorithm 1 is the whole set of every $W_j$, i.e.,

$$f(W_j) \geq f(\tilde{W}_j), \qquad \forall \tilde{W}_j \subset F_j. \qquad (8)$$

*Proof:* Algorithm 1 is equivalent to the process shown in Fig. 2. Each subcontract can be characterized by two values about the data quality: $D_{ij}$ and $D_{ij} + L(x_{ij})$. We represent each subcontract using a line segment, as shown in Fig. 2, with the two ends of the line segment assigned the two values $D_{ij}$ and $D_{ij} + L(x_{ij})$ on the axis of data quality, respectively. If we use a line perpendicular to the data quality axis and cross the axis at $L$, all those horizontal line segments can be categorized into three classes: class one includes those line segments that completely lie on the left side of $L$, class two includes those lie completely on the right side of $L$, and class three includes those intersecting with the vertical line $L$.

---

**Algorithm 1:** Quality-Driven Auction

Categorize $\{x_{ij}\}$ into corresponding $F_j$ ;
$W_j = \emptyset$ ;
$Q_2 = \emptyset$ ;
$G = G_{high} = G_{low} = 0$ ;
$i = 0$ ;
Sort data in $F_j$ according to their values of $D_{ij} + L(x_{ij})$ in the descending order ;
$G_{high} = \max_{F_j}\{D_{ij} + L(x_{ij})\}$ ;
Sort data in $F_j$ according to their values of $D_{ij}$ in the descending order ;
$G_{low} = \min_{F_j}\{D_{ij}\}$;
**while** *true* **do**
　　$G = (G_{low} + G_{high})/2$ ;
　　**for** $s = 1$ *to* $n$ **do**
　　　　**if** $D_{sj} > G$ **then**
　　　　　　$W_j \leftarrow W_j \cup x_{sj}$ ;
　　　　**else if** $D_{sj} + L(x_{sj}) > G$ **then**
　　　　　　$Q_2 \leftarrow Q_2 \cup x_{sj}$ ;
　　　　**end**
　　**end**
　　**if** $G > \sum_{W_j \cup Q_2} L(x_{ij})$ **then**
　　　　$G_{high} = G$ ;
　　**else if** $G < \sum_W L(x_{ij})$ **then**
　　　　$G_{low} = G$ ;
　　**else**
　　　　break ;
　　**end**
**end**
Find $V = \arg\max_{T \subset Q_2} (f(W_j \cup T))$ ;
**return** $W_j = W_j \cup V$;

---

It is straightforward that the three classes are in fact representing the three subcontract sets $Q_1$, $Q_2$, and $Q_3$ mentioned previously, respectively. As stated in Theorem 1, if we know the exact sum of the quality of the elements in the winner data set, it is safe to say that all the elements in $Q_3$ are in the winner set, whereas the ones in $Q_1$ are definitely not. The challenge is to find the value of such "exact sum." Fortunately, we could find the possible range of such value with the facilitation of Theorem 2 and the dichotomy in Algorithm 1.

We first randomly choose a value $G$, which separates all line segments into three classes. If the sum of the length of the line segments in $Q_2$ and $Q_3$ is less than $G$, the sum of length of the line segments in the winner data set will be on the left side of $G$ by the first item of Theorem 2. We will try to move $G$ to the left in this case. If the sum of the length of the line segments only in $Q_3$ is already greater than $G$, the sum of length of the line segments in the winner data set will be on the right side of $G$ by the second item of Theorem 2. We will try to move $G$ to the right in this case. In this way, we are able to narrow down the possible range of $L$ corresponding to the winner data set on the data quality axis by the third item of Theorem 2. All subcontracts completely on the right side of the range must be in the winner data set, and all those completely on the left side of the range must not. We now only need to search those subcontracts in $Q_2$, which can maximize the

system's social welfare. Those subcontracts are denoted to be in the set $V$ in Algorithm 1. Finally, the winner data set should be $W_j \cup V$. ∎

It may be noticed that it takes exponential computation complexity finding $V$; however, the searching range in Algorithm 1 is just $Q_2$, which is much smaller than it would be in the reverse Vickrey auction (all possible subcontracts combinations). Thus, Algorithm 1 shows much higher computational efficiency in practice as to be shown in the performance evaluation section. It may also be noticed that the computational complexity of Algorithm 1 is also related to $R(\cdot)$ that influences the distribution of $D_{ij}$. However, it is interesting to find that the computation complexity of Algorithm 1 is, in fact, more dependent on the characteristics of MCS data. This is because $D_{ij}$ is dependent on the quality of each data in the first hand, and the quality of submitted data will be normally diversified for the nature of MCS system, where there is no guarantee on the quality of workers. The diversity of workers makes the line segments along the data quality axis sparsely distributed, which reduces the size of $Q_2$ and thus reduces the computation complexity.

## V. PROVING PROPERTIES OF QUALITY-DRIVEN AUCTION

Here, we prove that QDA has the properties of individual rationality, truthfulness, platform profitability, and social welfare maximization. To prove the first property, we consider the following two situations: First, the worker claims his true cost as the lowest acceptable payment, and second, the worker claims an arbitrary price, where the corresponding winner data sets are $W$ and $W^*$, respectively.

*Lemma 3:* If the data $x_{st}$ is in both $W$ and $W^*$, then

$$W = W^*.$$

*Proof:* Since data $x_{st}$ is accepted in both sets, and all the other subcontracts never change, we need to examine if we can find a set of data excluding $x_{st}$, which can maximize the social welfare. In the platform's perspective, social welfare is $f_p(W) = R(\sum_W L(x_{ij})) - \sum_W b_{ij}$. We can regard $R(\sum_W L(x_{ij}))$ as $R(L(x_{st})) + R_\Delta$, where $R_\Delta$ stands for the marginal revenue of all the data except $x_{st}$ in the winner set. Since $R(L(x_{st})) - b_{st}$ is a constant when we know that $x_{st}$ must be in the winner set and its claimed price, no matter what the value of $b_{st}$ is, we need to find a set to maximize $R_\Delta - \sum_{W/\{x_{st}\}}$. Since this expression is independent of $x_{st}$, the result of finding such set will make no difference, which leads to $W = W^*$. The social welfare in the two cases could be different, but this does not mean that $b_{st}$ can be arbitrary large, or the data may not be accepted, which contradicts the condition of this lemma. ∎

*Lemma 4:* If the data $x_{st}$ is in both $W$ and $W^*$, then

$$f_p(W) = f_p(W^*) + b_{st} - k_{st}.$$

*Proof:* This is equally to prove

$$R\left(\sum_W L(x_{ij})\right) - \sum_{\frac{W}{\{x_{st}\}}} b_{ij} = R\left(\sum_{W^*} L(x_{ij})\right) - \sum_{W^*} b_{ij}.$$

According to Lemma 3, $W = W^*$ in this case, the result is straightforward. ∎

*Theorem 4:* QDA is truthful.

*Proof:* We consider the following two situations: First, the worker claims his true cost as the lowest acceptable payment; second, the worker claims an arbitrary price. If $x_{ij} \in W$ and $x_{ij} \in W^*$, we prove that the utilities in both cases are the same. If $x_{ij} \notin W$ and $x_{ij} \notin W^*$, the utilities are of course both zero.

With Lemma 4, the utility for the data with an arbitrary price is

$$
\begin{aligned}
u_{st}^* &= p_{st}^* - k_{st} = f_p(W^*) - f_p(W^{st}) + b_{st} - k_{st} \\
&= f_p(W) - f_p(W^{st}) \\
&= f_p(W) - f_p(W^{st}) + k_{st} - k_{st} \\
&= p_{st} - k_{st} \\
&= u_{st}.
\end{aligned}
$$

In our proof, $f_p(W^{*st}) = f_p(W^{st})$ because $x_{st}$ is in neither $W^{st}$ nor $W^{*st}$, which means that whatever the contract is will not affect the result of the winner set; thus, $W^{*st} = W^{st}$.

If $x_{ij} \in W$ but $x_{ij} \notin W^*$, the user will lose his chance to profit by claiming a price other than true cost. If $x_{ij} \notin W$ but $x_{ij} \in W^*$, for $b_{ij} > k_{ij}$, this will not happen because the lower the asked price is, the greater chance it is accepted. Then, if $b_{ij} < k_{ij}$, we prove that the payment for the data $p_{ij}$ will be even lower than its cost, i.e.,

$$
\begin{aligned}
p_{st}^* &= f_p(W^*) - f_p(W^{st}) + b_{st} > k_{st} \\
&\times f_p(W^*) - k_{st} + b_{st} > f_p(W^{st}) \\
R\left(\sum_{W^*} L(x_{ij})\right) &- \sum_{W^*} b_{ij} + b_{st} - k_{st} > f_p(W^{st}).
\end{aligned}
$$

In conclusion, if $b_{ij} > k_{ij}$, the data could be accepted or unaccepted, and the corresponding utility is $u_{ij}$ or zero, respectively. If the worker claims the true cost, the data will also have the two results, and the utility is the same. Consequently, the worker would rather claim the true cost to get more chance that his data are accepted. If $b_{ij} < k_{ij}$, however, there are three possible utilities for that data, which are $u_{ij}$, zero, or negative. Therefore, the worker will not claim $b_{ij} < k_{ij}$ to prevent loss. ∎

*Lemma 5:* If data $x_{st} \in W$, then

$$f_p(W) \geq f_p(W^{st}).$$

*Proof:* Since the winner data set is the set that can maximize the social welfare in the platform's perspective, if $f_p(W^{st})$ is greater than $f_p(W)$, then choosing $W^{st}$ will still be a better choice to maximize the social welfare even if data $x_{st}$ exists. This contradicts the fact that data $x_{st}$ is winner data; thus, $f_p(W) \geq f_p(W^{st})$. ∎

*Theorem 5:* QDA is individual rational.

*Proof:* If data $x_{st}$ is rejected, corresponding payment will be zero; thus its utility is zero. We only need to consider the case when $x_{st}$ is accepted. In the last theorem, we already proved that the user will only claim the true cost. Then, with Lemma 5, we have

$$
\begin{aligned}
u_{st} &= p_{st} - k_{st} \\
&= f_p(W) - f_p(W^{st}) + k_{st} - k_{st} \\
&= f_p(W) - f_p(W^{st}) \geq 0.
\end{aligned}
$$
∎

*Lemma 6:* If the data $x_{st}$ is in $W$, then $f_p(W^{st}) \geq f_p(W/\{x_{st}\})$.

*Proof:* The left-hand side is the social welfare when data $x_{st}$ is not in the winner set. To obtain $W^{st}$, the platform may add some other data to the winner set, although the social welfare will not be better than the original case according to Lemma 3. However, $f_p(W^{st})$ will be still larger than $f_p(W/\{x_{st}\})$, which simply deletes $x_{st}$ from the winner set. The process to get $W^{st}$ is to get $W/\{x_{st}\}$ first, meaning to find whether there are other data that can increase the social welfare if included. ∎

*Theorem 6:* QDA is platform profitable.

*Proof:*

$$u_p = R\left[\sum_W L(x_{ij})\right] - \sum_W p_{ij}$$

$$= R\left[\sum_W L(x_{ij})\right] - \sum_W \left\{f_p(W) + b_{ij} - f_p(W^{ij})\right\}$$

$$\geq \sum_{x_{st} \in W} \left\{R\left[\sum_W L(x_{ij})\right] - R\left[\sum_{\frac{W}{\{x_{st}\}}} L(x_{ij})\right]\right\}$$

$$- \sum_W \left\{f_p(W) + b_{ij} - f_p(W^{ij})\right\}$$

$$= \sum_{x_{st} \in W} \left\{R\left[\sum_W L(x_{ij})\right] - R\left[\sum_{\frac{W}{\{x_{st}\}}} L(x_{ij})\right]\right.$$

$$\left. - f_p(W) - b_{st} + f_p(W^{st})\right\}.$$

Since $R[\sum_W L(x_{ij})] - f_p(W) = \sum_W b_{ij}$, then

$$u_p = \sum_{x_{st} \in W} \left\{f_p(W^{st}) - R\left[\sum_{\frac{W}{\{x_{st}\}}} L(x_{ij})\right] + \sum_{\frac{W}{\{x_{st}\}}} b_{ij}\right\}$$

$$= \sum_{x_{st} \in W} \left\{f_p(W^{st}) - f_p\left(\frac{W}{\{x_{st}\}}\right)\right\} \geq 0.$$

∎

*Theorem 7:* QDA is social welfare maximal.

*Proof:* The QDA is truthful, thus maximizing $f_p(W)$ is equivalent to maximizing the sum of every player's utility in the game, including the platform. We can substitute every $f_p(W)$ with $f(W)$ in all formulas above. The social welfare optimal is important because if we take the users and the platform as a whole sensing system, then the social welfare function can be regarded as the efficiency function of the sensing network, i.e., the revenue of the accepted data, minus the cost spent on sensing. ∎

## VI. APPLYING QUALITY-DRIVEN AUCTION TO THE INDOOR LOCALIZATION SYSTEM

We apply the QDA-based incentive mechanism to an indoor localization system, where the worker needs to report his current location and corresponding Wi-Fi RSS fingerprint. The challenge is how to measure the reliability of the submitted fingerprints. We propose to transform the unreliability of the submitted data to the unreliability of human beings' positioning
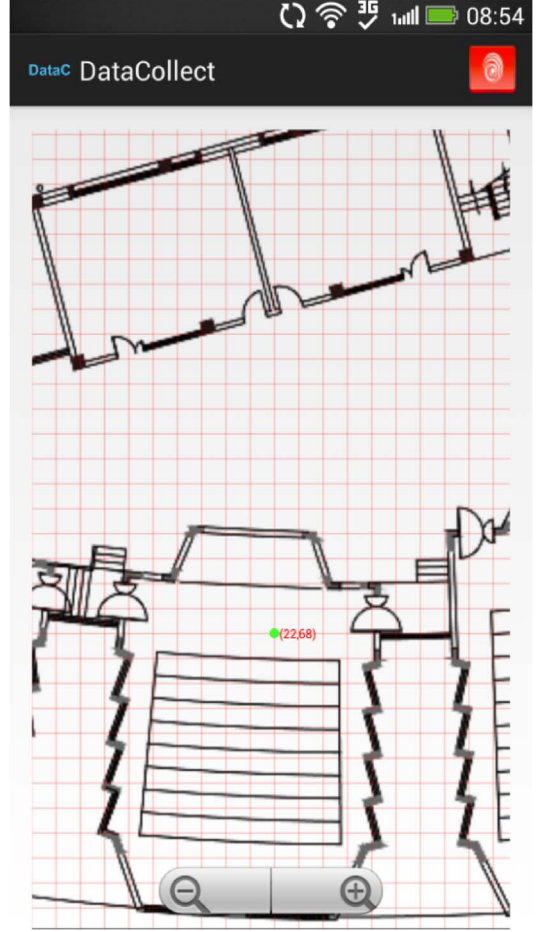


Fig. 3. Screenshot of the MCS App.

sense, which can be profiled by experiments performed in advance once and for all.

We develop an App for users who could be enrolled as workers. Fig. 3 shows a screenshot of the App, where the green spot is where a worker thinks he is standing, which is termed as *center* for the remainder of this paper. The coordinate (22, 68) denotes the estimated position by the worker, which is measured by counting the number of squares horizontally and vertically. Other eight cross points surrounding the center are termed as *neighbors*. The task released by the requester is to measure the fingerprint of the center. The worker stands on the place where he believes is the center, presses a button, and the corresponding fingerprint will be sent to a small-scale cloud implemented with CloudFoundry [25]. The measured fingerprint is the data to be submitted, whereas the corresponding cost of the task is autonomically computed by the App based on the resource of the mobile phone.

However, it is possible that the place the worker stands on is not the exact center the requester is interested in, which incurs error of the submitted data because of human beings' positioning sense error. Most likely, the worker actually chooses a neighbor around the center. Specifically, worker $i$ actually standing in the area $k$, thought himself in the area $j$, will submit data $x_{ij}$ that actually is $x_{ik}$. The requester needs to measure the fingerprints of different centers, which can be modeled as different types of tasks denoted as $M = \{1, 2, \ldots, m\}$.

We use $r_k$ to denote the probability that the data submitted for center $k$ are indeed measured on center $k$ and that $r_{kj}$ is the probability that the data are actually for center $k$ but that the worker thought it was for the center $j$. Since the two probabilities are closely related to the probability that human beings' positioning sense error occurs, they can be obtained by general-purpose experiments performed in advance. In this paper, the building where we perform the experiment for our system is divided into grid according to the layout of the ceramic tiles, which are widely used in Chinese buildings. Since the ceramic tile is usually in the shape of a square with an area of 1.2 m $\times$ 1.2 m, it needs reasonable efforts to find the specific center.

After the platform received the data $x_{ij}$, it actually regards the data as $x_{ia}$, where $a$ is the area that has the largest $r_k \cdot r_{kj}$. In particular, each fingerprint's requester could hold probability $\alpha$ as a threshold to benchmark $r_a \cdot r_{aj}$. Then, we can define the $L(x_{ij})$ in the indoor localization system as

$$L(x_{ij}) = \ln \left( \frac{r_a \cdot r_{aj}}{\alpha} \right)$$
$$a = \arg\max\{r_k \cdot r_{kj} | k \in M\}.$$

It is straightforward to see that $L(x_{ij}) < 0$ when $r_a \cdot r_{aj}$ is smaller than $\alpha$. The platform could reject the data since it is not very reliable. With the definition of data reliability, the QDA model can be applied to the indoor localization system.

## VII. Performance Evaluation

We perform our experiments in three classrooms and one corridor of the Dongzhong Yuan building at the Shanghai Jiao Tong University campus, which is 100 m$^2$ in size. More than 500 fingerprints have been collected with 20 mobile phones. The costs of the smart phones for performing the fingerprint collection are configured to be uniformly distributed over $[0, k_{\max}]$, which models resource levels of a large-scale crowd. We perform each experiment 100 times and take the average value as the result. We verify if the important properties of our proposed scheme indeed hold in practice and examine the corresponding cost in terms of the computational complexity.

### A. Truthfulness and Individual Rationality

We first verify the individual rationality and truthfulness of the incentive mechanism and show the results in Fig. 4. The figure shows that any worker may be unable to obtain higher utility by deviating from the true price, which is the cost incurred to collect the data. It also shows that any worker will obtain nonnegative utility if the true price is claimed.

In our experiment, the deviation is measured by the ratio of the claimed price to the true price. The claimed price is termed as the high price if the ratio of itself to the true price is greater than one, and it is termed as the low price if the corresponding ratio is less than 1. We ask 20 workers to sample 50 locations and let each worker report a random high and low price each for 100 times at each location, respectively. We randomly pick a worker and measure the utility obtained and plot the results in Fig. 4. It is obvious that asking a true price is always the best choice, which leads to nonnegative utility. We can see that the worker's utilities are different in different locations because the reliability of each sampling is different from others. The
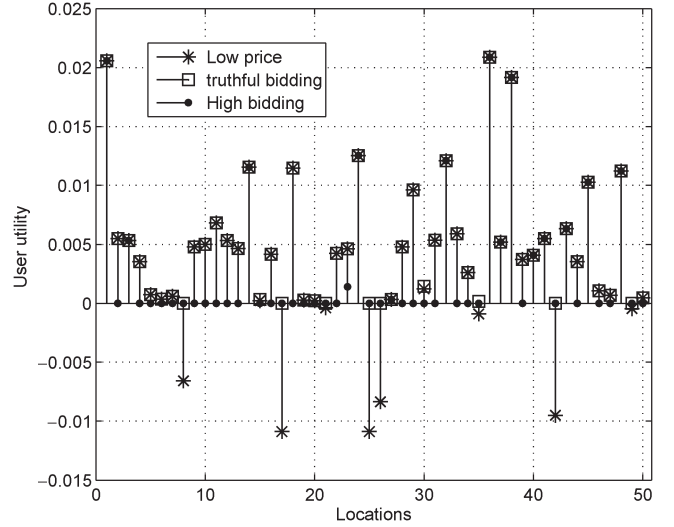


Fig. 4. Verification of truthfulness and individual rationality.
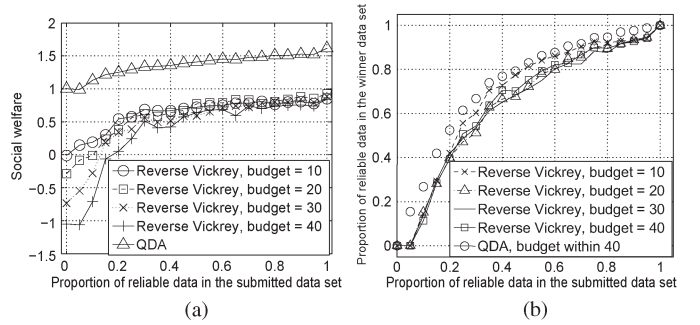


Fig. 5. (a) Social welfare and (b) effectiveness of quality discrimination.

utilities are almost zero in all locations when the high price is claimed; this is because the platform will exclude such workers from the candidates set, as described in Algorithm 1. The utility could be less than 0 in some cases when the low price is claimed; this is because the payoff cannot even balance off the cost with the low price.

### B. Social Welfare

The overall social welfare of the proposed scheme includes workers' utilities plus the platform's utility. The results are shown in Fig. 5(a), where the performance of our mechanism is compared with that of the traditional reverse Vickrey auction with different budgets. We implement a straightforward extension to the reverse Vickrey auction to let it take the platform's utility into consideration when calculating the social welfare, to make a fair comparison. Therefore, the difference between the social welfare is just because of the quality of the accepted data.

Our mechanism is quality driven, and the utilization of funding is more effective. In our mechanism, the platform has no need to consume up the budget, and only needs to regard the budget as an upper bound. In the experiments, the budget for our mechanism is set to be within 40, and the platform can spend the funding based on the reliability of the submitted data. It is shown in Fig. 5(a) that the social welfare of our scheme is always higher than that of the extended reverse Vickrey auction. With the quality driver, our scheme is always able to select data with high reliability and refuses to accept the submitted data

if all of them are unreliable. Given a data set with different proportions of reliable data, our scheme can always achieve higher social welfare. In contrast, the extended reverse Vickrey auction accepts all data without checking the corresponding contribution to the social welfare; therefore, the performance is lower than that of ours.

### C. Quality Discrimination

We examine how effectively our proposed scheme can discriminate data with different levels of reliability. We make different data sets from the entire database of the indoor localization system and configure the proportion of data with high probability for each set. We want to check if the data with high reliability can be selected by our scheme if the data set were submitted to the platform. The results are shown in Fig. 5(b), where our mechanism is also compared with the extended reverse Vickrey auction scheme with different budgets. The horizontal axis denotes the proportions of reliable data in the submitted data set, and the vertical axis denotes the proportions of reliable data in the resulted winner data set, which are selected by our proposed scheme. We also set the budget of QDA scheme as 40. It is obvious that the QDA can select more reliable data compared with the reverse Vickrey auction, which indicates the effectiveness of our scheme. It is interesting that the proportion of reliable data in the winner data set selected by the reverse Vickrey auction does not increase with the corresponding budget. This is because the more funding the platform has, the more unreliable data can be selected by the reverse Vickrey auction since it is not quality driven. The results corroborate the results earlier. Since more reliable data are selected by our proposed scheme, the corresponding social welfare is higher under the QDA.

### D. Computational Cost

Here, the computational cost of the QDA is evaluated with respect to three factors: the number of workers, the reliability of workers, and the cost of workers for performing a task.

The computational cost with regard to the number of recruited workers is closely related to the scalability of any MCS incentive mechanism. Ideally, the computational cost of the platform should be independent of the number of recruited workers. The proposed QDA scheme provides a smart way to avoid searching the entire $2^n$ contracts, where $n$ is the number of possible contract space. Fig. 6 presents the computation time that it takes for the QDA to find a winner data set in comparison with that for the reverse Vickrey auction. It is obvious that the proposed QDA outperforms the reference scheme. Since the reverse Vickrey auction needs to search all possible contracts to determine the winner data set, it takes more than 1 min for the platform to finish the calculation when there are only 30 workers. It is easy to see that the proposed QDA is very computational efficient, with computation time negligible.

The computation cost of the platform is also related to the quality of the submitted data and the cost of workers for performing sensing tasks. As described in Section VI, the reliability of the submitted data for the indoor localization system under study is measured by $L(x_{ij})$. Fig. 7(a) shows
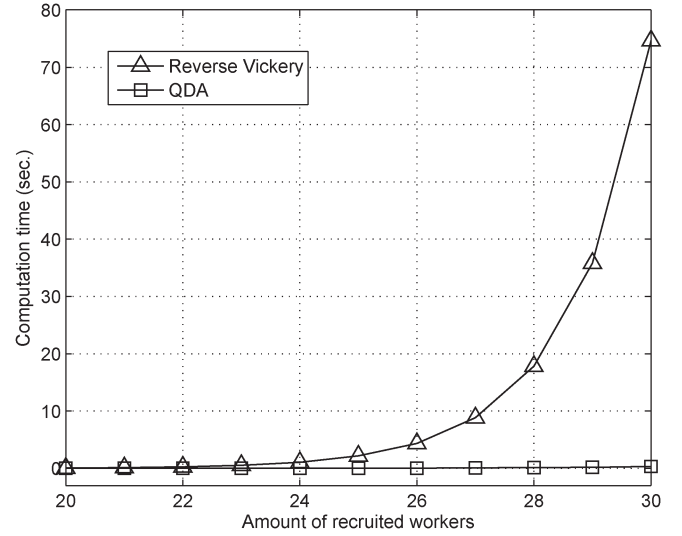


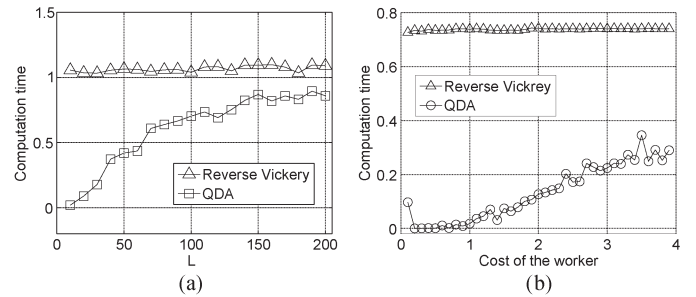Fig. 6.   Computation time with respect to the number of recruited workers.



Fig. 7.   Computation time with respect to the (a) data reliability and (b) worker's cost.

how long it takes to obtain the winner data set with submitted data of different levels of reliability. It is easy to see that the proposed QDA takes less time than the reverse Vickrey auction. The reason is the same as above: The searching space is narrowed. It is obvious that the QDA takes longer time to find the winner data set when the reliability of submitted data increases, whereas the reverse Vickrey auction makes almost no difference. This is because the QDA is sensitive to the data reliability, whereas the reverse Vickrey auction does not take the reliability into consideration.

The costs of workers to perform tasks will also influence the computation cost. This is because when the price is extremely low, buying any data can bring marginal utility to the platform, which makes it hard to determine a winner data set. On the other hand, the higher the sensing cost is, the higher payment the platform needs to give. Even if the quality of data is not considered, it is still a challenge to find out a winner data set and keep the platform profitable at the same time. This is why the computation cost increases for both the QDA and the reverse Vickrey auction as shown in Fig. 7(b). However, the computation cost of the QDA is still less than that of the reverse Vickrey auction.

### VIII. CONCLUSION

In this paper, we have proposed an incentive mechanism based on QDA. The mechanism is specifically for the MCS system, where the worker is paid off based on the quality of

sensed data instead of working time, as adopted in the literature. We have theoretically proved that the mechanism is truthful, individual rational, platform profitable, efficient, and social-welfare optimal. Moreover, we have incorporated our incentive mechanism into a Wi-Fi fingerprint-based indoor localization system, to incentivize the MCS-based fingerprint collection. We have presented a probabilistic scheme to evaluate the accuracy of the data submitted, which is to resolve the issue that the ground truth for the data accuracy is unavailable. We have realized and deployed the indoor localization system to evaluate our proposed incentive mechanism and presented extensive experimental results.

## References

[1] N. Lane *et al.*, "A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 140–150, Sep. 2010.

[2] "Mobile Millennium," Berkeley, CA, USA, 2011. [Online]. Available: http://traffic.berkeley.edu

[3] M. Mun *et al.*, "PIER, the personal environmental impact report, as a platform for participatory sensing systems research," in *Proc. ACM MobiSys*, 2009, pp. 55–68.

[4] R. Rana, C. Chou, S. Kanhere, N. Bulusu, and W. Hu, "Earphone: An end-to-end participatory urban noise mapping," in *Proc. ACM/IEEE IPSN*, 2010, pp. 105–116.

[5] V. Conitzer and T. Sandholm, "Revenue failures and collusion in combinatorial auctions and exchanges with VCG payments," in *Proc. 6th AAMAS Int. Conf. Agent-Mediated Electron. Commerce: Theories Eng. Distrib. Mech. Syst.*, 2004, pp. 1–14.

[6] L. Gao, Y. Xu, and X. Wang, "MAP: Multiauctioneer progressive auction for dynamic spectrum access," *IEEE Trans. Mobile Comput.*, vol. 10, no. 8, pp. 1144–1161, Aug. 2011.

[7] L. Gao and X. Wang, "A game approach for multi-channel allocation in multi-hop wireless networks," in *Proc. ACM MobiHoc*, 2008, pp. 303–312.

[8] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. ACM Mobicom*, 2012, pp. 195–205.

[9] L. Duan *et al.*, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE INFOCOM*, 2012, pp. 1701–1709.

[10] D. Zhao, X. Li, and H. Ma, *OMG: How Much Should I Pay Bob in Truthful Online Mobile Crowdsourced Sensing?* [Online]. Available: http://arxiv.org/abs/1306.5677

[11] M. Bernstein, J. Brandt, R. Miller, and D. Karger, "Crowds in two seconds: Enabling realtime crowd-powered interfaces," in *Proc. 24th Annu. ACM Symp. User Interface Softw. Technol.*, 2011, pp. 33–42.

[12] P. Minder, S. Seuken, A. Bernstein, and M. Zollinger, "CrowdManager—Combinatorial allocation and pricing of crowdsourcing tasks with time constraints," in *Proc. 2nd Workshop Social Comput. User Gener. Content*, Valencia, Spain, Jun. 7, 2012, pp. 1–18.

[13] D. R. Karger, S. Oh, and D. Shah, "Efficient crowdsourcing for multi-class labeling," in *Proc. ACM SIGMETRICS*, 2013, pp. 81–92.

[14] S. He, D. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *Proc. IEEE INFOCOM*, 2014, pp. 745–753.

[15] Z. Jiang *et al.*, "Communicating is crowdsourcing: Wi-Fi indoor localization with csi-based speed estimation," *J. Compumt. Sci. Technol.*, vol. 29, no. 4, pp. 589–604, Jul. 2014. [Online]. Available: http://arxiv.org/abs/1307.6349

[16] H. Liu *et al.*, "Push the limit of wifi based localization for smartphones," in *Proc. ACM MobiCom*, 2012, pp. 305–316.

[17] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, "Centaur: Locating devices in an office environment," in *Proc. ACM MobiCom*, 2012, pp. 281–292.

[18] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proc. ACM MobiCom*, 2012, pp. 293–304.

[19] S. Sen, R. R. Choudhury, and S. Nelakuditi, "Spinloc: Spin once to know your location," in *Proc. ACM HotMobile*, 2012, Art. ID. 281.

[20] Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. ACM MobiCom*, 2012, pp. 269–280.

[21] L. Cong and W. Zhuang, "Non-line-of-sight error mitigation in mobile location," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 560–573, Mar. 2005.

[22] K. Chen and W. Tsai, "Vision-based autonomous vehicle guidance for indoor security patrolling by a SIFT-based vehicle-localization technique," *IEEE Trans. Veh. Technol.*, vol. 59, no. 7, pp. 3261–3271, Sep. 2010.

[23] S. Coleri *et al.*, "RSSI-fingerprinting-based mobile phone localization with route constraints," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 423–428, Jan. 2014.

[24] L. Liao *et al.*, "Two birds with one stone: Wireless access point deployment for both coverage and localization," *IEEE Trans. Veh. Technol.*, vol. 60, no. 5, pp. 2239–2252, Jan. 2011.

[25] E. Dudin and Y. G. Smetanin, "A review of cloud computing," *Sci. Tech. Inf. Process.*, vol. 38, no. 4, pp. 280–284, 2011.
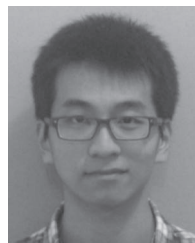
**Yutian Wen** is currently working toward the B.E. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China.

His research interests include crowdsourcing-based indoor localization.

**Jinyu Shi** (S'12) is currently working toward the B.E. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China. Her research interests include crowdsourcing and green data centers.

Ms. Shi is one of the students from the IEEE Class.

**Qi Zhang** is currently working toward the B.E. degree with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China.
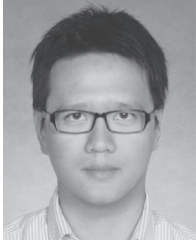
His current research interests include mechanism design for crowdsourcing/crowdsensing systems.

**Xiaohua Tian** (S'07–M'11) received the B.E. and M.E. degrees in communication engineering from Northwestern Polytechnical University, Xian, China, in 2003 and 2006, respectively, and the Ph.D. degree from Illinois Institute of Technology, Chicago, IL, USA, in 2010.

Since June 2013, he has been with the Department of Electronic Engineering, Shanghai Jiao Tong University, as an Assistant Professor with the title of SMC-B scholar.

Dr. Tian served/will serve as a Technical Program Committee (TPC) member for the 2014–2015 IEEE Conference on Computer Communications (IEEE INFOCOM); the 2011–2015 IEEE Global Communications Conference Symposium on Wireless Networking, Cognitive Radio Networks, and Ad Hoc and Sensor Networks; and for the 2013 and 2015 IEEE International Conference on Communications Symposium on Ad Hoc and Sensor Networks and Next Generation Networking. He is best demo/poster award committee member for the 2014 IEEE INFOCOM; a TPC Chair for the IEEE International Conference on Communications in China (ICCC) International Workshop on Internet of Things and the 2014 International Conference on Wireless Algorithms, Systems and Applications Symposium on Next Generation Networking; and a Local Management Chair for IEEE ICCC 2014. He serves as an editorial board member on the computer science subsection of the journal *SpringerPlus*, and as a Guest Editor for the *International Journal of Sensor Networks*.

**Zhengyong Huang** received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2012. He is currently working toward the M.S. degree with Shanghai Jiao Tong University.

His main research interests include indoor localization.

**Hui Yu** received the B.S. degree from Tong Ji University, Shanghai, China, in 1992 and the M.S. degree from Shanghai Jiao Tong University, Shanghai, in 1997.

He is currently a Senior Engineer with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include mobile communications, software radio and cognitive radio, channel coding, and modulation for wireless communications.

**Yu Cheng** (S'01–M'04–SM'09) received the B.E. and M.E. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1995 and 1998, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2003.

From September 2004 to July 2006, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON. Since August 2006, he has been with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, USA, where he is currently an Associate Professor. His research interests include next-generation Internet architectures and management, wireless network performance analysis, network security, and wireless/wireline interworking.

Dr. Cheng served as a Cochair for the Wireless Networking Symposium of the 2009 IEEE International Conference on Communications (ICC); the Communications QoS, Reliability, and Modeling Symposium of the 2011 IEEE Global Communications Conference (GLOBECOM); the Signal Processing for Communications Symposium of IEEE ICC 2012 and the Ad Hoc and Sensor Networking Symposium of IEEE GLOBECOM 2013; and as a Technical Program Committee Cochair for the 2011 International Conference on Wireless Algorithms, Systems, and Applications and the 2015 International Conference on Computing, Networking, and Communications. He is a founding Vice Chair of the IEEE Communication Society Technical Subcommittee on Green Communications and Computing. He serves as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and the New Books and Multimedia Column Editor for *IEEE Network*. He received the Best Paper Award at the 2007 International Conference on Heterogeneous Networking for Quality, Reliability, Security, and Robustness and the 2011 IEEE ICC; the National Science Foundation CAREER Award in 2011; the IIT Sigma Xi Research Award in the junior faculty division in 2013; and the Best Paper Runner-up Award from the 2014 Association for Computing Machinery International Symposium on Mobile Ad Hoc Networking and Computing.

**Xuemin (Sherman) Shen** (M'97–SM'02–F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982 and the M.Sc. and Ph.D. degrees from Rutgers University, Piscataway, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering.

From 2004 to 2008, he was the Associate Chair for Graduate Studies with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, where he is currently a Professor and the University Research Chair. He is a coauthor/editor of 15 books and more than 800 papers and book chapters in wireless communications and networks, control, and filtering. His research interests include resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks.

Dr. Shen served as the Technical Program Committee Chair/Cochair for the 2014 IEEE International Conference on Computer Communications and 2010 Fall IEEE Vehicular Technology Conference (VTC); as the Symposia Chair for the 2010 IEEE International Conference on Communication; as the Tutorial Chair for 2011 Spring IEEE VTC and 2008 IEEE ICC; as the Technical Program Committee Chair for the 2007 IEEE Global Communications Conference; as the General Cochair for the 2015 Association for Computing Machinery International Symposium on Mobile Ad Hoc Networking and Computing, the 2007 International Conference on Communication and Networking in China, and the 2006 International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks; and as the Chair for IEEE Communications Society Technical Committee on Wireless Communications and peer-to-peer Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE NETWORK, *Peer-to-Peer Networking and Application*, and *IET Communications*; as a Founding Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; as an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, *Computer Networks*, *ACM Wireless Networks*, etc.; and as a Guest Editor for IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS MAGAZINE, *ACM Mobile Networks and Applications*, etc. He received the Excellent Graduate Supervision Award in 2006 and the Outstanding Performance Award in 2004, 2007, and 2010 from the University of Waterloo; the Premiers Research Excellence Award in 2003 from the Province of Ontario, Canada; and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. He is an elected member of IEEE Communication Society Board of Governors and the Chair of Distinguished Lecturers Selection Committee, a Distinguished Lecturer of IEEE Vehicular Technology and Communications Societies, and a Fellow of the Engineering Institute of Canada and Canadian Academy of Engineering. He is a registered Professional Engineer of Ontario.