Quality-Driven Auction based Incentive Mechanism for Mobile Crowd Sensing

Yutian Wen, Jinyu Shi, 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 are proposed, it is still an open issue 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. 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, in order to incentivize the MCS based fingerprints collection. We present a probabilistic model to evaluate the reliability of the submitted data, which is to resolve 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, mobile crowd sensing, indoor localization.**

I. Introduction

Mobile phones are increasingly intelligent in past years, which not only own the processing power comparable to that of laptops, but also accommodate a rich set of sensors such as accelerometer, compass, gyroscope, GPS, microphone and camera. With appropriate organization, mobile phones could form sensing networks enabling new mobile applications across various domains [1]. For example, GPSes in mobile phones can be utilized to collect traffic information and help users estimate travel time [2]; Phone sensors can help tracking the individual behavior to evaluate the impact on the environment pollution [3]; Phone-embedded microphones can help create noise maps in different areas [4].

Employing sensors embedded in mobile phones to collect data presents a new sensing paradigm known as *mobile crowd sensing* (MCS), which is different from the traditional sensing techniques relying on static sensors such as wireless sensor networks. The MCS system basically consists of the mobile phone users acting as sensing service providers (workers), 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 [9], because 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], [6]. The Stackelberg game, contract theory, auction theory are employed to model the interactions between workers and the platform [7]–[9], [11]. 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, especially 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 non-trivial [12], 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; 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 (LBS) [19], where collecting the received signal strength (RSS) of Wi-Fi access points (APs) within buildings is an important part. 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 [14], [16], [17]. Although researchers mentioned or indicated the concept of incorporating the MCS into the RSS collecting process, most of work still focuses on the localization technique itself, and fundamentals and important details of the incentive mechanism design are still unclear.

Yutian Wen, Jinyu Shi, Qi Zhang, Xiaohua Tian, Zhengyong Huang and Hui Yu are with Electronic Engineering Department, Shanghai Jiao Tong University, Shanghai, China, 200240. E-mail: {nogerw, mizaoyu, qzhang, xtian, zhengyonghuang, yuhui}@sjtu.edu.cn. Xiaohua Tian is the corresponding author.

Yu Cheng is with the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, Illinois, USA 60616. E-mail: cheng@iit.edu.

Xuemin (Sherman) Shen is with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo Ontario N2L 3G1 Canada. E-mail: xshen@bbcr.uwaterloo.ca.

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. 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 which can maximize the social welfare of the system. We show that the proposed mechanism can encourage more submissions of high-quality data with a lower computational complexity. In our mechanism, the platform has no need to know the cost of each individual worker, which is supposed to be the 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 MCSed 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 for all. The profile gives the probability of a human's incorrectness of locality, which is then utilized to find the submitted data with 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 $100m^2$ in a building of 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 [20]. Extensive experimental results are presented to illustrate the performance of our scheme.

The remainder of this paper is organized as follows. Section II gives a more detailed overview of related work. Section III presents the system model and design challenges. Section IV elaborates the incentive mechanism design. Section V proves important properties of the mechanism. Section VI describes how the incentive mechanism is incorporated into the indoor localization system. Section VI presents the experimental results. Section VIII gives the conclusion remarks.

II. Related Work

A. General-Purpose Incentive Mechanisms

Models in game theory can be borrowed to design the incentive mechanism. Yang et al. propose two types of incentive mechanisms for the MCS system in the perspectives of the agent platform and mobile users, respectively [7]. 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. classify the MCS system into two classes: data acquisition and distributed computing [8]. 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 Stckelberg 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. propose an online incentive mechanism for the case where workers arrive one by one [9], 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.

In order to shorten the crowd response time, Bernstein et al. propose 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 [10]. Based on the retainer model, Patrick et al. propose a combinatorial allocation and pricing scheme for crowdsourcing tasks with time constraints [11]. 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 (VCG) based rule.

Although refering to the reverse Vickrey auction model, our scheme considers the reliability of the submitted data, which provides a 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 [15]–[17], [19]. The fingerprint-based localization technique can be divided into the training and localizing phase. In the training phase, the fingerprints at positions of interests are collected into a database by measuring the RSS of Wi-Fi APs around. In the localizing phase, the system will search the users 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 just well-trained experts; therefore, offloading the RSS collection to the MCS becomes a consensus [14], [17], [19] but barely investigated.

The challenge of designing an incentive mechanism for MCSing 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 non-trival, because there is no perfect benchmark. Karger et al. study 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 [12]. 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. propose a pricing mechanism based on bargaining theory for mobile crowd sensing, which realistically considers the task performing cost and the market demand [13]. Our work however focuses on the influence of data quality on the payoffs of the data collection workers.

TABLE I
MAIN NOTATIONS

n	Number of interested workers
N	Interested workers set
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

III. System Model and Design Challenges

A. System Model

We consider the MCS system consists 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 process below will be performed for each time slot.

- Contract determination;
- Winner data set determination;
- Payment determination;
- Response and update.

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

Contract determination: After receiving the demand, a set of workers $N = \{1, 2, ..., n\}$ 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 $M = \{1, 2, ..., m\}$ to denote all types of data need 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 as one task, and there are totally m tasks here. Each pair of data and claimed price is termed as a *sub-contract* denoted as c_{ij}, and all sub-contracts 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.

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 $W_j = argmax\{f(W_j)|W_j \in F_j\}$, 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} 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. In order to incentivize workers to submit high-quality data, we use k_{ij} to denote the cost of user i if submitted data x_{ij} is accepted by a single requester, and use $l \cdot k_{ij}$ to denote the cost if x_{ij} is 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 they have submitted and 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.

Sine workers are paid off according to the quality of submitted data, the proposed model can encourage higher-quality participants 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 achieve higher effectiveness.

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 each candidate set will bring. Secondly, 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 each can bring, and choose the candidate that can bring the greatest social welfare as 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 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 with 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 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 a high computation 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: 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 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.

In the following sections, we are to propose a new mechanism to resolve all the issues above, which is termed as *Quality-Driven Auction*.

IV. Quality-Driven Auction

A. Overview

The idea of the Quality-Driven Auction (QDA) is as follows. First, calculate a particular value for each sub-contract, which reflects the extent to which the data is worth of buying and sort all sub-contracts 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:

$$u_i = \sum_{j \in M_i^*} u_{ij} = \sum_{j \in M_i^*} p_{ij} - \sum_{j \in M_i^*} k_{ij} \geq 0, \tag{1}$$

Truthfulness: No worker can achieve a better utility by submitting a 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}:

$$\begin{aligned} u_{ij}(c_{ij}, c_{-ij}) &= u_{ij}((b_{ij}, x_{ij}), c_{-ij}) \\ &\leq u_{ij}((k_{ij}, x_{ij}), c_{-ij}), \end{aligned} \tag{2}$$

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 0.

$$u_p = R(\sum_W L(x_{ij})) - \sum_W p_{ij}, \tag{3}$$

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 a reliable data ($L(x_{ij}) > 0$) into the winner data set will always bring the platform benefit. With more and more reliable data accepted, the marginal revenue brought by a new data will become less and less. Consequently the platform has a decreasing marginal revenue.

Social welfare maximization: The total payoffs across all players is 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:

$$\begin{aligned} f(W) &= \sum_{i\in N} u_i + u_p \\ &= \sum_{W}(p_{ij} - k_{ij}) + R(\sum_{W} L(x_{ij})) - \sum_{W} p_{ij} \\ &= R(\sum_{W} L(x_{ij})) - \sum_{W} k_{ij}. \end{aligned} \tag{4}$$

We use W^{st} to represent the winner data set where data x_{st} is definitely rejected:

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

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:

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

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} does to the whole system. However, if a data is not accepted, then it's payment will be 0.

It is worth mentioning that there may exist more than one winner sets, that is, $\exists W_1$, $\exists W_2$ 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 violate 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 choose 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 0 no matter the data is selected or not.

B. Particular Value of the Sub-Contract

The first step of the QDA is to calculate the particular value of each sub-contract mentioned earlier and sort all sub-contracts by the value. This value is a measurement that to what extent the data is worthy to buy, 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 the revenue can be obtained by the platform. The curve stands for $R(\cdot)$, which is the revenue function. For a 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

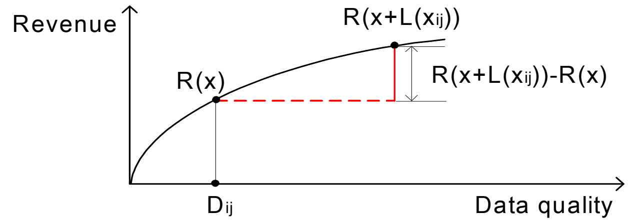


Fig. 1. Illustration of D_{ij}

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 will finally less than 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 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 a 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, a data with larger D_{ij} is more worthy to buy.

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

Lemma 1: For $\forall H \subset F, x_{ij} \in H$, 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 a larger welfare than the former one, that means data x_{ij} actually makes the social welfare lower. ∎

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

Proof: Because $R'()$ is monotonically decreasing, adding data x_{ij} to a set with smaller total value will have a higher marginal revenue while the cost remains the same, which will lead to a higher social welfare. Consequently, when adding x_{ij} into a set whose total value is D_{ij} will not decrease the social welfare, adding it to a set with smaller total value will have an 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:

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

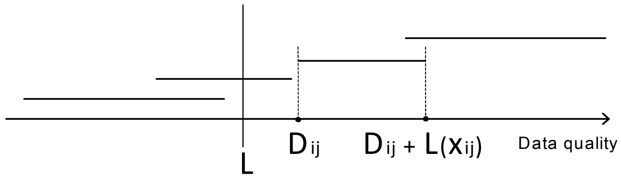


Fig. 2. Illustration of algorithm 1.

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

$$Q_3(G) = \{x_{ij} | G < D_{ij}\}$$

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: $Q_1(L) \cap W = \emptyset$, $Q_3(L) \subseteq W$.

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 $Q_3(L) \subseteq W$. ■

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 contradict 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 $W \cap Q_1(G) = \emptyset$, so $W \subseteq Q_3(G)$. Consequently, $\sum_{Q_3(G)} L(x_{ij}) \geq L > G$, a contradiction. Therefore, if L is larger than G, then $W \cap Q_2 \neq \emptyset$. In order 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 $\sum_{Q_3(L)} L(x_{ij}) < L < \sum_{Q_2(L) \cup Q_3(L)} L(x_{ij})$, so $W \cap Q_2(G) \neq \emptyset$.

■

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 thus omitted here.

Theorem 3: The output of Algorithm 1 is the whole set of every W_j,

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

Proof: Algorithm 1 is equivalent to the process illustrated in Fig. 2. Each sub-contract can be characterized by two values

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 $W_j = W_j \cup V$;

about the data quality: D_{ij} and $D_{ij} + L(x_{ij})$. We represent each sub-contract 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 on the left side of L, class two includes those that completely on the right side of L, and class three includes those that are intersecting with the vertical line L.

It is straightforward that the three classes are in fact representing the three sub-contract sets Q_1, Q_2 and Q_3 mentioned earlier, 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 while 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 sub-contracts 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 sub-contracts in Q_2, which can maximize the system's social welfare. Those sub-contracts 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 sub-contracts 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 MCSed 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 reduce the computation complexity.

V. Proving Properties of QDA

In this section, 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 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 sub-contracts 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_Δ 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

$$\begin{aligned} R(\sum_W L(x_{ij})) - &\sum_{W/\{x_{st}\}} b_{ij} \\ &= R(\sum_{W^*} L(x_{ij})) - \sum_{W^*} b_{ij} \end{aligned}$$

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

Theorem 4: Quality-Driven Auction is truthful.

Proof: We consider the following two situations: first, the worker claims his true cost as 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 0.

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.

$$\begin{aligned} p^*_{st} = f_p(W^*) - f_p(W^{st}) + b_{st} &> k_{st} \\ f_p(W^*) - k_{st} + b_{st} &> f_p(W^{st}) \\ R(\sum_{W^*} L(x_{ij})) - \sum_{W^*} b_{ij} + b_{st} - k_{st} &> f_p(W^{st}) \end{aligned}$$

In conclusion, if $b_{ij} > k_{ij}$, the data could be accepted or unaccepted, and the corresponding utility is u_{ij} or 0, 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}, 0 or negative. Therefore, the worker will not claim $b_{ij} < k_{ij}$ to prevent loss.

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

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

Proof: Since the winner data set is the set which 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 a winner data, thus $f_p(W) \geq f_p(W^{st})$. ■

Theorem 5: Quality-Driven Auction is individual rational.

Proof: If data x_{st} is rejected, corresponding payment will be 0, thus its utility is 0. We only need to consider the case when x_{st} gets accepted. In last theorem, we already proved that the user will only claim the true cost. Then, with Lemma 5,

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

■

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

Proof: The LHS 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 which can increase the social welfare if included. ■

Theorem 6: Quality-driven Auction is platform profitable.

Proof:

$$\begin{aligned} u_p &= R[\sum_W L(x_{ij})] - \sum_W p_{ij} \\ &= R[\sum_W L(x_{ij})] - \sum_W \{f_p(W) + b_{ij} - f_p(W^{ij})\} \\ &\geq \sum_{x_{st} \in W} \{R[\sum_W L(x_{ij})] - R[\sum_{W/\{x_{st}\}} L(x_{ij})]\} \\ &- \sum_W \{f_p(W) + b_{ij} - f_p(W^{ij})\} \\ &= \sum_{x_{st} \in W} \{R[\sum_W L(x_{ij})] - R[\sum_{W/\{x_{st}\}} L(x_{ij})] \\ &- f_p(W) - b_{st} + f_p(W^{st})\}; \end{aligned}$$

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

$$\begin{aligned} u_p &= \sum_{x_{st} \in W} \{f_p(W^{st}) - R[\sum_{W/\{x_{st}\}} L(x_{ij})] + \sum_{W/\{x_{st}\}} b_{ij}\} \\ &= \sum_{x_{st} \in W} \{f_p(W^{st}) - f_p(W/\{x_{st}\})\} \geq 0. \end{aligned}$$

■

Theorem 7: Quality-Driven Auction is social welfare maximal.

Proof: The Quality-Driven Auction is truthful, thus maximizing $f_p(W)$ is equivalent to maximizing the sum of every

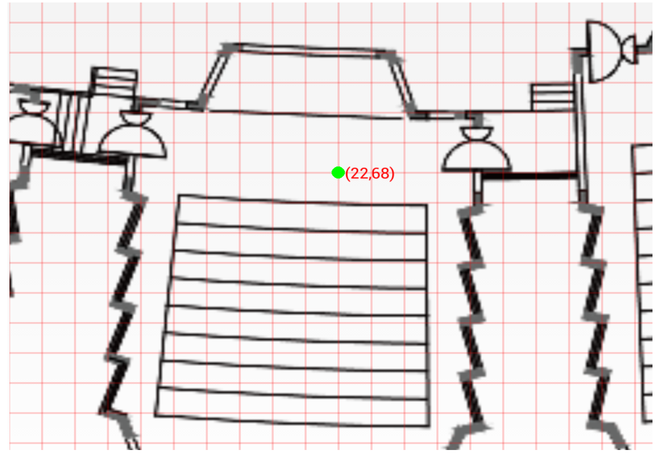


Fig. 3. Screenshot of the MCS App.

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

We develop an App for users who could be enrolled as workers. Figure 3 illustrates a screenshot of the App, where the green spot is where a worker think he is standing, which is termed as *center* for the rest of the 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 8 cross points surrounding the center are termed as *neighbours*. 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 [20]. The measured fingerprint is the data to be submitted, while 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 choose a neighbor around the center. Specifically, a worker i actually standing in the area k, thought himself in the area j, will submit the 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 task denoted as $M = \{1, 2, ..., m\}$.

We use r_k to denote the probability that the data submitted for center k is indeed measured on center k, and r_{kj} is the probability that the data is actually for center k but the worker thought it is 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 our study, 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 $1.2m \times 1.2m$-square, 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 $r_k \cdot r_{kj}$. Particularly, each fingerprints requester could hold a probability α as a threshold, in order to benchmark $r_a \cdot r_{aj}$. Then we can define the $L(x_{ij})$ in the indoor localization system as

$$L(x_{ij}) = ln(\frac{r_a \cdot r_{aj}}{\alpha}),$$
$$\alpha = argmax\{r_k \cdot r_{kj} | k \in M\}$$

It is straightforward to see that $L(x_{ij}) < 0$ when $r_a \cdot r_{aj}$ is smaller than α. 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 3 classrooms and 1 corridor of Dongzhong Yuan building in Shanghai Jiao Tong University campus, with $100m^2$ in size. More than 500 fingerprints are 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 is to model resource levels of 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 is unable to obtain a 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 a non-negative 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 high price if the ratio of itself to the true price is greater than 1, and it is termed as low price if the corresponding ratio is less than 1. We ask 20 workers to sample 50 locations and let each worker to report a random high and low price each for 100 times at each location, respectively. We randomly

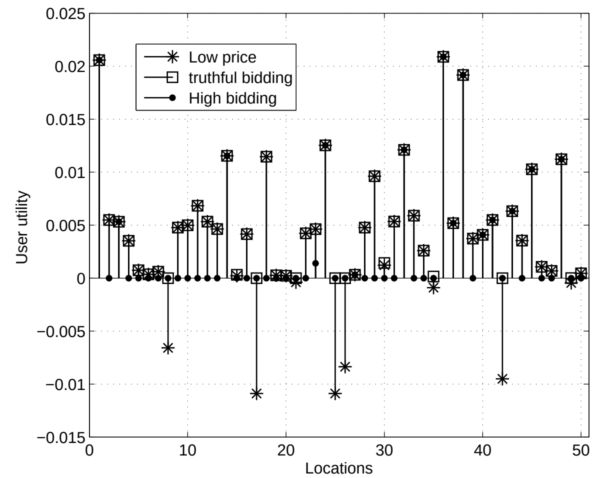


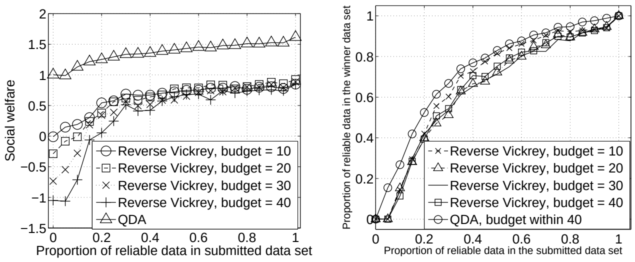
Fig. 4. Verification of truthfulness and individual rationality.

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 a non-negative 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 utilities are almost 0 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 can not 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 illustrated 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, in order 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 extended reverse Vickrey auction. With the quality driver, our scheme is always able to select data with high reliability, and refuse 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 a higher social welfare. In contrast, the extended reverse Vickrey auction accepts all data



(a) Social welfare.

(b) Effectiveness of quality discrimination.

Fig. 5. Social welfare and effectiveness of quality discrimination.

without checking the corresponding contribution to the social welfare, therefore the performance is lower than that of ours.

C. Quality Discrimination

We examine how effective 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 in the section above. Since more reliable data are selected by our proposed scheme, the corresponding social welfare are higher under the QDA.

D. Computational Cost

This subsection evaluates the computational cost of the QDA 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. Figure 6 presents the computation time it takes for the QDA to find a winner data set in comparison with that for the reverse Vickrey

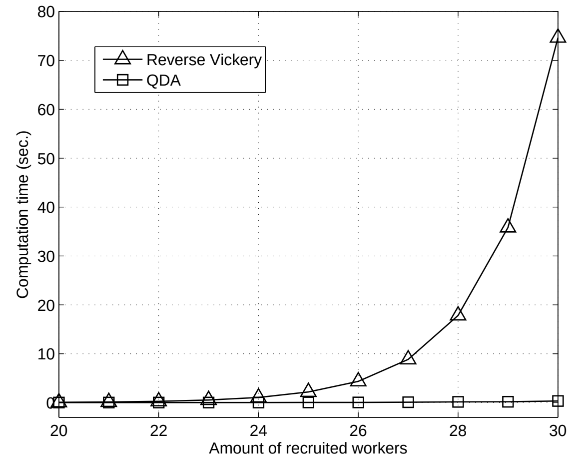
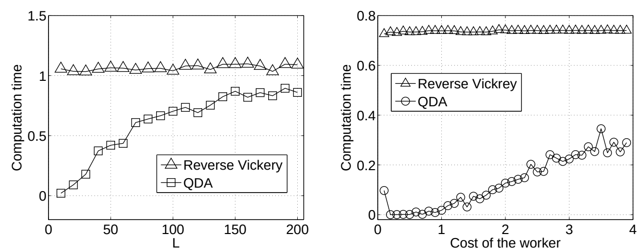


Fig. 6. Computation time with respect to the number of recruited workers.



(a) Computation time with respect to the submitted data reliability.

(b) Computation time with respect to the worker's cost.

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

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 minute 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 $L(x_{ij})$. Figure 7 (a) illustrates how long 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, while the reverse Vickrey auction makes almost no difference. This is because the QDA is sensitive to the data reliability, while 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 a 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. Conclusions

In this paper, we have proposed an incentive mechanism based on a quality-driven auction. 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, in order to incentivize the MCS based fingerprints 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, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, “A survey of mobile phone sensing,” *IEEE Communication Magazine* , vol. 48, no. 9, pp. 140–150, Sept. 2010.

[2] UC Berkeley/Nokia/NAVTEQ, “Mobile Millennium,” http://traffic.berkeley.edu

[3] M. Mun, S. Reddy, K. Shilton, N. Yau, J. Burke, D. Estrin, M. Hansen, E. Howard, R. West, and P. Boda, “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. The 6th AAMAS international conference on Agent-Mediated Electronic Commerce: theories for and Engineering of Distributed Mechanisms and Systems*, pp. 1-14, 2004.

[6] L. Gao, Y. Xu and X. Wang, “ MAP: Multiauctioneer progressive auction for dynamic spectrum access,” *IEEE Transactions on Mobile Computing*, vol. 10, no. 8, pp. 1144-1161, Aug. 2011.

[7] 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.

[8] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa and J. Walrand, “Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing,” in *Proc. IEEE INFOCOM*, 2012, pp. 1701–1709.

[9] D. Zhao, X. Li and H. Ma, “OMG: How much should I pay Bob in truthful online mobile crowdsourced sensing?” Jun. 24, 2013, Online: http://arxiv.org/abs/1306.5677.

[10] M. Bernstein, J. Brandt, R. Miller and D. Karger, “ Crowds in two seconds: Enabling realtime crowd-powered interfaces,” in *The 24th Annual ACM Symposium on User Interface Software and Technology*, 2011, pp. 33–42.

[11] P. Minder, S. Seuken, A. Bernstein and M. Zollinger, “CrowdManager - Combinatorial Allocation and Pricing of Crowdsourcing Tasks with Time Constraints,” in *The 2nd Workshop on Social Computing and User Generated Content*, Jun. 7, 2012, Valencia, Spain.

[12] D. R. Karger, S. Oh, D. Shah, “Efficient crowdsourcing for multi-class labeling,” in *Proc. ACM SIGMETRICS*, 2013.

[13] 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.

[14] Z. Jiang, J. Zhao, X. Li, W. Xi, K. Zhao, S. Tang and J. Han “Communicating is crowdsourcing: Wi-Fi indoor localization with CSI-based speed estimation,” Jul. 24, 2013, Online: http://arxiv.org/abs/1307.6349.

[15] H. Liu., Y. Gan, J. Yang., S. Sidhom, Y. Wang, Y. Chen, and F. Ye, “Push the limit of wifi based localization for smartphones,” in *Proc. ACM Mobicom*, 2012.

[16] R. Nandakumar, K. K. Chintalapudi, and V. N. Padmanabhan, “Centaur: locating devices in an office environment,” in *Proc. ACM Mobicom*, 2012.

[17] A.Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, “Zee: zero-effort crowdsourcing for indoor localization,” in *Proc. ACM Mobicom*, 2012.

[18] S. Sen, R. R. Choudhury and S. Nelakuditi, “Spinloc: spin once to know your location,” in *Proc. ACM HotMobile*, 2012.

[19] Z. Yang, C. Wu and Y. Liu, “Locating in fingerprint space: Wireless indoor localization with little human intervention,” in *Proc. ACM Mobicom*, 2012.

[20] E. Dudin and Y. G. Smetanin, “A review of cloud computing,,” *Scientific and Technical Information Processing*, vol. 38, no. 4, pp. 280–284, 2011.