Incentive Mobile Crowd Sensing for Time and Data Oriented Tasks

Project Report of Qi Zhang(5110309556)
Yutian Wen, Qi Zhang(5110309556),
Advised by Prof.Xiaohua Tian, Prof. Xinbing Wang

***Abstract*—Mobile phone crowdsourcing is a new paradigm which takes advantage of the pervasive smartphones for data collection or distributed computation. A critical issue for the paradigm is to incentivize users to provide services with their smartphones. While some incentive mechanisms are proposed in past years, it is still an open issue how to recruit more participating smartphones. In order to achieve the full potential of the crowdsourcing network, the mechanism is also expected to have certain properties like truthfulness and individual rationality. Prior approaches have been done for such a mechanism, however, they either cannot work properly for both data collection system and distributed computation system, or miss some of the important properties for this paradigm. In this paper, we propose an incentive mechanism working well for both time oriented and data oriented system. 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 mobile crowd sensing(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.**

I. Introduction

Mobile phones are increasingly intelligent in past years, which not only possess powerful and power-efficient processors, outstanding battery life, abundant memory, but also accommodate a rich set of sensors such as accelerometer, compass, gyroscope, GPS, microphone and camera. With appropriate organization, mobile phones could form collaboration system 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].

The mobile phone collaboration system basically consists of the mobile phone users acting as service providers (workers), requesters who want to get service from users, and an agent platform acting as a medium to recruit workers for requesters. Duan et al. classify the mobile phone collaboration applications into two categories: data acquisition and distributed computing [6].

In data acquisition applications, a requester wants to acquire enough data from smartphone users to build up a database. For example, an Android phone collects its location data via GPS every few seconds and transmits the data to Google. The phone also transmits back the name, location, and received signal strength(RSS) of any nearby Wi-Fi networks. After collecting enough location data from users, Google can successfully build a massive database capable of providing location-based services. 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. Besides phone resources, the user will potentially incur privacy leakage as his cost when performing the sensing task. We denote this kind of system as *data oriented* system.

In distributed computing applications, a requester wants to solve complex engineering or commercial problems inexpensively using the computing power of mobile phones. Since millions of smartphones remain unused most of the time, the requester might want to solicit smartphone collaborations. In this case, a user's collaboration cost may be due to loss of energy and reduction of physical storage. Thus, users' workload is measured by the working time. We denote this kind of system as *time oriented* system.

Most of the existing platforms recruit volunteers as workers [7]. However, both time oriented tasks and data oriented tasks will consume workers' phone resources and potentially incur privacy leakage. Consequently, designing a proper incentive mechanism for the mobile phone collaboration system is vitally important.

Game theory is used to address the issue because of its straightforward suitability. The Stackelberg game, contract theory, auction theory are employed to model the interactions between workers and the platform [5]–[7], [9]. While these models can be theoretically proved having favored characteristics such as truthfulness and profitability, applying the theory to the practice is hardly straightforward. In the time oriented tasks, each worker has an maximum workload and the payment to the worker is determined by his actual working time. In the data oriented tasks, workers' payment should be determined by the quality of the sensing data he uploads. The challenge for the practical incentive mechanism design for the mobile phone collaboration system is threefold: 1) the appropriate theoretical framework to model the actual interaction between workers and the platform is still unavailable; 2)In time oriented tasks, how to determine the workload effectively is still an open question;3) the effective approach to evaluate the quality of the crowd-sensed data needs more investigation.

Specifically, our contributions are as follows:

- We design an incentive mechanism for data oriented system (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 as a contract. We present an effective algorithm to select a group of data which can maximize the social welfare, and prove that this algorithm is much more computational efficient than those adopted in the literature. 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.
- We apply the quality-driven auction to the time oriented system. Similarly, every time a customer wants to campaign a time oriented task like crowd computing, he sends a request to the platform. The users who are interested in a particular task respond with the minimum payment he would accept, and the maximum time he is willing to spend for the task. The time oriented mechanism retain the desired properties in its data oriented counterpart.We prove that the mechanism is truthful, individual rational, platform profitable, and social-welfare optimal.

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 data oriented incentive mechanism design. Section V elaborates the time oriented incentive mechanism design. Section VI 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 [5]. 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. classifies the MCS system into two classes: data acquisition and distributed computing [6]. 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, which is inappropriate to be applied in practice.

B. Incentive Mechanisms for Specific Purposes

Zhao et al. propose an online incentive mechanism for the case where workers arrive one by one [7], 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 [8]. Based on the retainer model, Patrick et al. propose a combinatorial allocation and pricing scheme for crowdsourcing tasks with time constraints [9]. 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 referring 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.

III. System Model

A. Framework

We consider the system consisting of three kinds of players: workers, agent platform and requesters. We model the interaction between the players in both time oriented and data oriented systems. Specifically, the interactions below will take place on the platform.

- Contract collection;
- Winner contract determination;
- Payment determination;
- Response and update

Contract collection: A set of n workers, denoted by $N = \{1, 2, ..., n\}$, are willing to perform the task requested on the platform. The task is either data oriented or time oriented. To complete for the task, each worker i replies the platform with a *contract*. We use c_i^d and c_i^t respectively to denote worker i's contract in data and time oriented system.

In the data oriented system, $c_i^d = \{b_i^d, x_i^d\}$, where x_i^d is the data that worker i collects for the task. Bid b_i^d is worker i's lowest acceptable payment for collecting x_i^d. Besides, we assume that each worker has a cost k_i^d for collecting the data.

In the time oriented system, we assume each worker has a capacity T_i, which is the maximum time he is able to perform the task. A time oriented contract c_i^t consists of worker's bid b_i^t, which is his unit time acceptable payment, and the maximal time worker i is willing to spend on the task, denoted as t_i, i.e. $c_i^t = \{b_i^t, t_i\}$. Also, each worker has a unit time cost k_i^t for performing the time oriented task.

Winner determination: After collecting all the contracts, the platform needs to determine winner workers $W \subseteq N$. In the data oriented system, we denote winners' data set as W^d, which will be accepted by requesters and get the payment. In the time oriented system, the platform needs to determine the actual work time, i.e. workload, for each worker. We use $x_i^t \in [0, t_i]$ to denote the workload for worker i's contract c_i^t, which indicates that x_i^t should be less than the maximal work time t_i. We use $X^t = \{x_1^t, x_2^t, ..., x_n^t\}$ to denote how the platform allocate workload.

Payment determination: After determine the winners, the platform needs to calculate payment the requester should pay. If worker i is a winner, his payment should be more than his bid . Otherwise, payment is equal to 0. Also, we denote the payment in data oriented system as p_i^d and p_i^t is the payment per unit time in the time oriented system.

B. Design Goals

The solution to the winner determination problem(WDP) and payment determination problem(PDP) should be truthful, individual rational, platform profitable, and social welfare optimal. We give the definitions of these desired properties, which are slightly different between data and time oriented systems.

Individual rationality: We use u_i^d and u_i^t respectively to differentiate worker i's utility in data and time oriented system. Individual rationality means that winner workers will have a utility greater than 0, i.e.

$$u_i^d = p_i^d - k_i^d \geq 0 \tag{1}$$

$$u_i^t = x_i^t (p_i^t - k_i^t) \geq 0 \tag{2}$$

Truthfulness: In the data oriented contract, the worker truthfully set the bid as his true cost of data collection. No worker can achieve a better utility by submitting a lowest acceptable payment other than his cost, i.e.

$$\begin{aligned} u_i^d(c_i^d, c_{-i}^d) &= u_i^d((b_i^d, x_i^d), c_{-i}^d) \\ &\leq u_i^d((k_i^d, x_i^d), c_{-i}^d), \end{aligned} \tag{3}$$

where c_{-i}^d is the set contracts of workers excluding i.

In the time oriented system, workers will truthfully submit the cost as his bid, as well as the capacity as his maximum working time. Workers will have the largest utility by submitting this truthful strategy.

$$\begin{aligned} u_i^t(c_i^t, c_{-i}^t) &= u_i^t((b_i^t, t_i), c_{-i}^t) \\ &\leq u_i^t((k_i^t, T_i), c_{-i}^t), \end{aligned} \tag{4}$$

Platform profitability: The utility of the platform is greater than 0.

$$u_p^d = R(\sum_W L^d(x_i^d)) - \sum_W p_i^d \geq 0, \tag{5}$$

$$u_p^t = R(\sum_W L^t(x_i^t)) - \sum_W p_i^t x_i^t \geq 0, \tag{6}$$

where $L^d(x_i^d)$ is the evaluation of the quality of data x_i^d, and $L^t(x_i^t)$ is the evaluation of worker i's workload in the time oriented task. We may consider $R(\cdot)$ is the revenue function with the following properties:

$$R(0) = 0, R'(x) > 0, R''(x) < 0, \tag{7}$$

which indicates that the function has a decreasing marginal revenue.

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

All players include both workers and the platform, in contrary to most of the work in the literature, which only focuses either of them. We use the social welfare function $f^d(W)$, $f^t(W)$ to quantify the social welfare:

$$\begin{aligned} f^d(W) &= \sum_W u_i^d + u_p^d \\ &= \sum_W (p_i^d - k_i^d) + R(\sum_W L^d(x_i^d)) - \sum_W p_i^d \\ &= R(\sum_W L^d(x_i^d)) - \sum_W k_i^d. \end{aligned} \tag{8}$$

$$\begin{aligned} f^t(W) &= \sum_W u_i^t + u_p^t \\ &= \sum_W x_i^t(p_i^t - k_i^t) + R(\sum_W L^t(x_i^t)) - \sum_W p_i^t x_i^t \\ &= R(\sum_W L^t(x_i^t)) - \sum_W k_i^t x_i^t. \end{aligned} \tag{9}$$

IV. Data Oriented System

A. Overview

The idea of applying Quality-Driven Auction to solve WDP and PDP for data oriented system is as follows. First, calculate a particular value for each contract, which reflects the extent to which the data is worth of buying and sort all 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.

B. Winner Determination

The first step of winner determination for data oriented system is to calculate a particular value of each contract and sort all contracts by the value. Specifically, we use D_i to denote the value. It is defined as the largest $\sum_{s \neq i} L^d(x_s^d)$ with x_s^d in S_i, where S_i is a subset of the entire data set F and adding x_i^d to S_i will not cause the social welfare to decrease.

Formally, if $R(L^d(x_i^d)) \geq k_i^d$, then

$$\begin{aligned} D_i &= max\{D_i | R(D_i + L^d(x_i^d)) - \sum_{S_i \cup \{x_i^d\}} k_i^d \\ &\geq R(D_i) - \sum_{S_i} k_i^d\} \\ &= max\{D_i | R(D_i + L^d(x_i^d)) - R(D_i) - k_i^d \geq 0\}; \end{aligned}$$

If $R(L^d(x_i^d)) < k_i$, $D_i = 0$, which means that the revenue data x_i^d can bring to the platform is even lower than its own cost, adding data x_i^d to any set will make the social welfare decrease.

According to the definition above, if a data x_i^d has a larger D_i, the total value of the data that can be put into the winner set before x_i^d is selected is larger. Since the revenue function $R(\cdot)$ is monotonically increasing and adding x_i^d will not attenuate the social welfare, the platform can achieve higher social welfare. Therefore, a data with larger D_i is more worthy to buy.

To find such D_i for every x_i^d, we have no need to find the set S_i for every data. The D_i is actually just a number that can be found by the revenue function $R(\cdot)$. Let the x-axis be the sum of $L^d(x_i^d)$ and the y-axis be the revenue R, then we can find the largest number D_i along the x-axis while adding data x_i^d from that point's increment of R is larger than or at least equal to its own cost. Because $R''() < 0$, only x_i^d with $D_i > 0$ is possible to be selected into W as long as there exits at least one $D_i > 0$. Hence we first filter all $D_i = 0$ and assume that none of the x_i^d has $D_i = 0$ in the rest of the paper.

The value of the contract has many attributes that can be used in the following description.

Lemma 1: For $\forall H \subset F, x_i^d \in H$, if $\sum_H L^d(x_i^d) > D_i + L^d(x_i^d)$, then $f(H/x_i^d) > f(H)$.

Proof: According to the definition, $D_i + L^d(x_i^d)$ 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_i^d actually make the social welfare lower, because otherwise $H/\{x_i^d\}$ will be the set S in D_i. ■

Lemma 2: For $\forall H \subset F, x_i^d \notin H$, if $\sum_H (L^d(x_i^d)) < D_i$, then $f^d(H \cup x_i^d) > f^d(H)$.

Proof: Because $R'()$ is monotonically decreasing, adding data x_i^d 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_i^d into a set whose total value is D_i 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^d is a winner data set, and let $L^d = \sum_{W^d} L^d(x_i^d)$ be the total value of data in the winner data set. $\forall\ x_i^d \in F$, $\forall G \in R$, we divide the data set F into three sets:

$$Q_1(G) = \{x_i^d | D_i + L^d(x_i^d) < G\}$$

$$Q_2(G) = \{x_i^d | D_i < G < D_i + L^d(x_i^d)\}$$

$$Q_3(G) = \{x_i^d | G < D_i\}$$

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

Proof: According to lemma 1, if $D_i + L^d(x_i^d) < L^d$, then $x_i^d \notin W^d$ and $Q_1(L) \cap W^d = \emptyset$; according to lemma 2, if $D_i > L^d$, then $x_i^d \in W^d$ and $Q_3(L) \subseteq W^d$. ■

Theorem 2:

- If $G > \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$, then $L^d < G$;
- If $G < \sum_{Q_3(G)} L^d(x_i^d)$, then $L^d > G$;
- If $\sum_{Q_3(G)} L^d(x_i^d) < G < \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$, then $min\{D_i | x_i^d \in Q_2(G)\} < L^d < max\{D_i + L^d(x_i^d) | x_i^d \in Q_2(G)\}$

Proof:

- If $G > \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$ and $L^d \geq G$, then $L^d > \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$. Because W^d is the winner data set, $W^d \subseteq Q_2(L^d) \cup Q_3(L^d)$. Because $L^d > G$, $Q_2(L^d) \cup Q_3(L^d) \subseteq Q_2(G) \cup Q_3(G)$. $L^d = \sum_{W^d} L^d(x_i^d) \leq \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d) < G$, which is contradict to $L^d \geq G$;
- Similar to the proof above;
- We assume that $\sum_{Q_3(G)} L^d(x_i^d) < G < \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$. If L is larger than G, and $W^d \cap Q_2(G) = \emptyset$, apparently $W^d \cap Q_1(G) = \emptyset$, so $W^d \subseteq Q_3(G)$. Consequently, $\sum_{Q_3(G)} L^d(x_i^d) \geq L^d > G$, a contradiction. Therefore, if L is larger than G, then $W^d \cap Q_2 \neq \emptyset$. In order to keep at least one element of $Q_2(G)$ in W^d, there must exist at least one element with $D_i \in Q_2(G)$ with $L^d < D_i + L^d(x_i^d)$, thus $L^d < max\{D_i + L^d(x_i^d) | x_i^d \in Q_2(G)\}$. The proof for the case when L^d is smaller than G is likewise. If $L^d = G$, then $\sum_{Q_3(L^d)} L^d(x_i^d) < L^d < \sum_{Q_2(L^d) \cup Q_3(L^d)} L^d(x_i^d)$, so $W^d \cap Q_2(G) \neq \emptyset$.

■

We present the algorithm of determining the winner data set for a specific task for the convenience of presentation. The process of determining the entire submitted data set is similar thus omitted here. We explain the main idea of the algorithm as follows:

1) Obtain D_i for every data x_i^d;
2) Sort all the D_i, select a number G, and separate all these data into three sets: $Q_1(G)$, $Q_2(G)$ and $Q_3(G)$;
3) Add all the data into $Q_3(G)$ and abandon all the data in $Q_1(G)$.
4) If $G > \sum_{Q_2(G) \cup Q_3(G)} L^d(x_i^d)$ or $G < \sum_{Q_3(G)} L^d(x_i^d)$, update the value of G to narrow down the searching scope and return to step 4);
5) Otherwise, search all the possible candidates of being accepted for data in $Q_2(G)$;
6) Calculate the social welfare each candidate can bring including those data already in the winner set;

Algorithm 1: Data Oriented Winner Determination

```
W^d = ∅ ;
Q_2 = ∅ ;
G = G_high = G_low = 0 ;
i = 0 ;
Sort data according to their values of D_i + L^d(x_i^d) in the descending order ;
G_high = the highest D_i + L^d(x_i^d) ;
G_low = the lowest D_i;
while true do
    G = (G_low + G_high)/2 ;
    for s = 1 to n do
        if D_s > G then
            W^d ← W^d ∪ x_s^d ;
        else if D_s + L^d(x_s^d) > G then
            Q_2 ← Q_2 ∪ x_s^d ;
        end
    end
    if G > Σ_{W^d ∪ Q_2} L^d(x_i^d) then
        G_high = G ;
    else if G < Σ_{W^d} L^d(x_i^d) then
        G_low = G ;
    else
        break ;
    end
end
W^d = argmax_{W^d ∪ T, T ⊂ Q_2} f(W^d) ;
return W^d;
```

C. Payment Determination

We use W^s to represent the winner set when data x_s^d from worker s is definitely rejected, the corresponding social welfare can be written as:

$$f^d(W^s) = R(\sum_{W^s} L^d(x_i^d)) - \sum_{W^s} k_i^d. \tag{10}$$

Note that the actual cost is only known by the worker himself, thus the platform simply treats the lowest acceptable payment b_i^d as the cost for sensing data x_i^d. Consequently, the social welfare in the platform's perspective is:

$$f_p^d(W) = R(\sum_{W} L^d(x_i^d)) - \sum_{W} b_i^d \tag{11}$$

If worker $s \in W$, then his payment will be

$$p_s^d = f_p^d(W) - f_p^d(W^s) + b_s^d, \tag{12}$$

meaning that the incremental contribution data x_s^d does to the whole system. However, if a data is not accepted, then it's payment will be 0.

It is worth mentioning that 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, $f^d(W_1) = f^d(W_2) \geq f^d(W)$.

All these data are acceptable to the platform, and none of them violate the rule of payment. Although it may seem to be unfair to those who are ruled out of the winner set, it is easy to prove that choosing any one of those winner sets will not hinder the truthfulness and individual rationality. 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 $x_i^d \in W_2, x_i^d \notin W_1$, $p_i^d = f^d(W_1) - f^d(W^i) + b_i^d$. Now that $x_i^d \in W_2, x_i^d \notin W_1$, $f^d(W_1) = f^d(W^i)$, $p_i^d = b_i^d$. This means that all users will only claim $b_i^d = k_i^d$, and the utility for the data x_i^d is always 0 no matter the data is selected or not.

D. Proving Properties

We prove that the solution to WDP and PDP has the properties of truthfulness, individual rationality,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 winner set are W and W^* and the corresponding winner data set are W^d and W^{d*}, respectively.

Lemma 3: If the data x_s^d is in both W^d and W^{d*}, then

$$W^d = W^{d*}.$$

Proof: Since data x_s^d is accepted in both sets and all the other contracts never change, we need to examine if we can find a set of data other than x_s^d, which can maximize the social welfare. In the platform's perspective, social welfare is $f_p^d(W) = R(\sum_W L^d(x_i^d)) - \sum_W b_i^d$. We can regard $R(\sum_W L^d(x_i^d))$ as $R(L^d(x_s^d)) + R_\Delta$, where R_Δ stands for the marginal revenue of all the data except x_s^d in the winner set. Because $R(L^d(x_s^d)) - b_s^d$ is a constant when we know that x_s^d must be in the winner set and its claimed price, no matter what the value of b_s^d is, we need to find a set to maximize $R_\Delta - \sum_{W/\{s\}}$. Since this expression is independent of x_s^d, 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_s^d can be arbitrary large, or the data may not be accepted, which does not meet the condition of this lemma. ■

Lemma 4: If the data x_s^d is in both W^d and W^{d*}, then

$$f_p^d(W) = f_p^d(W^*) + b_s^d - k_s^d.$$

Proof: This is equally to prove

$$R(\sum_{W} L^d(x_i^d)) - \sum_{W/\{s\}} b_i^d = R(\sum_{W^*/\{s\}} L(x_i^d)) - \sum_{W^*/\{s\}} b_i^d$$

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 $x_i^d \in W^d$ and $x_i^d \in W^{d*}$, we prove that the utilities in both cases are the same. If $x_i^d \notin W^d$ and $x_i^d \notin W^{d*}$, the utilities are of course both 0. With Lemma 4, the utility for the data with an arbitrary price is

$$\begin{aligned} u_s^{*d} &= p_s^{*d} - k_s = f_p(W^*) - f_p(W^s) + b_s - k_s \\ &= f_p^d(W) - f_p^d(W^s) \\ &= f_p^d(W) - f_p^d(W^s) + k_s^d - k_s^d \\ &= p_s^d - k_s^d \\ &= u_s^d. \end{aligned}$$

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

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

$$\begin{aligned} p_s^{*d} = f_p^d(W^*) - f_p^d(W^s) + b_s^d &> k_s^d \\ f_p^d(W^*) - k_s^d + b_s^d &> f_p^d(W^s) \\ R(\sum_{W^*} L^d(x_i^d)) - \sum_{W^*} b_i^d + b_s^d - k_s^d &> f_p^d(W^s) \end{aligned}$$

In conclusion, if $b_i^d > k_i^d$, the data could be accepted and unaccepted, and the corresponding utility is u_i^d 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 $b_i^d < k_i^d$, however, there are three possible utilities for that data, which are u_i^d, 0 or negative. Therefore, the worker will not claim $b_i^d < k_i^d$ to prevent loss.

Lemma 5: If a data $x_s^d \in W^d$, then

$$f_p^d(W) \geq f_p^d(W^s).$$

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

Theorem 5: Quality-Driven Auction is individual rational.

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

$$\begin{aligned} u_s^d &= p_s^d - k_s^d \\ &= f_p^d(W) - f_p^d(W^s) + k_s^d - k_s^d \\ &= f_p^d(W) - f_p^d(W^s) \geq 0 \end{aligned}$$

■

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

Proof: The LHS is the social welfare when data x_s^d is not in the winner set. To obtain W^s, 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, $f_p^d(W^s)$ will be still larger than $f_p^d(W/\{s\})$, which simply deletes x_s^d from the winner set. The process to get W^s is to get $W/\{s\}$ first, meaning to find whether there are other data which can increase the social welfare if included. ■

Theorem 6: Quality-Driven Auction is social welfare maximal.

Proof: Because Quality-Driven Auction is truthful, maximizing $f_p^d(W)$ is equivalent to maximize the sum of every player's utility in the game, including the platform. Thus we can substitute every $f_p^d(W)$ with $f^d(W)$ 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. ■

V. Time Oriented System

A. Design Overview

In this section, we solve the winner determination and payment determination for the time oriented system. The mechanism in time oriented system also has the properties of individual rationality, truthfulness, platform profitability and social welfare maximization.

An important observation of time oriented task is that despite the difference between the efficiency of the participants, the workload is usually interchangeable among peers. So we assume the value that crowd create is a function of the linear superposition of each participants work, i.e.

$$R(\sum L_i^t(x_i^t)) = R(\sum e_i x_i^t)$$

We utilize thus e_i to determine the winners.

B. Winner Determination in Time Oriented System

We use $X^t = \{x_1^t, x_2^t, ...x_n^t\}$ to notate how the platform allocates workload. Further, we use $\mathcal{X}^t$ to notate the strategy space if X^t,

$$\mathcal{X}^t = \{(x_1^t, x_2^t, ..., x_n^t) | 0 \leq x_i^t \leq t_i\}$$

We give workload determination of time oriented system in Algorithm 2 and prove it is social welfare optimal.

Theorem: The work allocation determined by Algorithm 2 is the social optimal work allocation, i.e. $f^t(X^t) \geq f^t(\hat{X}^t), \forall \hat{X}^t \in \mathcal{X}^t$

Proof: Assume there exits $\hat{X}^t \neq X^t, s.t. f^t(X^t) \geq f^t(\hat{X}^t)$. Let $\sum e_i x_i^t$ be L^t, $\sum e_i \hat{x}_i^t$ be $\hat{L}^t$. The corresponding winner set are W and $\hat{W}$. If $\hat{L}^t > L^t$, then $\frac{dR(x)}{dx}|_{x=\hat{L}^t} < \frac{dR(x)}{dx}|_{x=L^t} = \frac{b_k^t}{e_k}$, b_k^t and e_k are the unit bid and work efficiency of the last user n_k in algorithm 2 to be added into the winner set. Because $\hat{L}^t > L^t$, there must be at least $\hat{L}^t - L^t$ unit of work

Algorithm 2: Workload Determination

$W=\emptyset, sum=0, cost=0, i=1$; $X^t=\{0,0,...,0\}$;
sort N according to $\frac{b_i^t}{e_i}$ in ascending order;
while $\frac{dR(x)}{dx}|_{x=sum} > \frac{b_i^t}{e_i}$ *and* $W \neq N$ **do**
  $W \leftarrow W \cup \{n_i\}$;
  if $\frac{dR(x)}{dx}|_{x=sum+t_i} \geq \frac{b_i^t}{e_i}$ **then**
    $x_i^t = t_i$;
    $sum \leftarrow sum + e_i t_i$;
    $cost \leftarrow cost + x_i^t b_i^t$;
    $i = i+1$;
  else
    $x_i^t = \frac{1}{\frac{dR^{-1}(y)}{dy}|_{y=\frac{b_i^t}{e_i}}} - sum$; $sum \leftarrow sum + e_i t_i$;
    $cost \leftarrow cost + x_i^t b_i^t$;
    break;
end
$k=i$;
$f^t(X) = R(sum) - cost$;
return $W, X^t, f^t(X)$;

allocated to a set of participants $T, T \subset \hat{W}$. All the member in W except are fully loaded, so $T \cap W \subseteq u_k^t$ Since $R(\cdot)$ has decreasing marginal value,

$$R(\hat{L}^t - RL^t) < \frac{dR(x)}{dx}|_{x=L^t}(\hat{L}^t - L^t) = \frac{b_k^t(L^t - \hat{L}^t)}{e_k}$$

Because $< \frac{dR(x)}{dx}|_{x=L^t} = \frac{b_k^t}{e_k}$

$$f^t(\hat{W}/T) - f^t(\hat{W}) = R(L^t) - R(\hat{L}^t) + \sum_{n_i \in T} t_i b_i^t \quad (13)$$

$$\geq \frac{b_k^t(L^t - \hat{L}^t)}{e_k} - \frac{b_k^t(L^t - \hat{L}^t)}{e_k} = 0 \quad (14)$$

So the social welfare will increase if the platform reject those contracts offered by users in T or reduce the extra burden on u_k^t. Proof for cases when $\hat{L}^t < L^t$ is omitted because of similarity. If $\hat{L}^t = L^t$, then $R(\hat{L}^t) = R(L^t)$. It is obvious that the social cost $\sum_{n_i \in W} t_i b_i^t$ is minimum, so $f^t(W)$ is still no less than $f^t(\hat{W})$. Therefore we proved the work allocation computed by Algorithm 2 is social optimal.

C. Payment Determination in Time Oriented System

Winner workers determined by Algorithm 2 will get paid for performing time oriented tasks. Let $\mathcal{X}^{ti}$ be the strategy space where worker i is rejected from the winners, i.e.

$$\mathcal{X}^{ti} = \{X^t | X^t \in \mathcal{X}^t, x_i^t = 0\}$$

We run Algorithm 2 on $\mathcal{X}^{ti}$ to get the workload X^{ti}. The platform will pay worker i

$$p_i^t x_i^t = f_p^t(X^t) + b_i^t x_i^t - f_p^t(X^{ti})$$

where $f_p^t(X^t)$ is the social welfare in the platform's perspective,

$$f_p^t(X^t) = R(\sum_W L^t(x_i^t)) - \sum_W b_i^t x_i^t.$$

Because $\mathcal{X}^{ti} \subseteq \mathcal{X}^t$,it is obvious that $f_p^t(X^t) \geq f_p^t(X^{ti})$

We claim that the payment determination is individual rational, truthful and platform profitable. We omit the proofs here since they are similar to the data oriented counterparts.

VI. Conclusions

In this paper, we have proposed an incentive mechanism for both time oriented and data oriented crowdsourcing system. The mechanism for time oriented tasks reaches social welfare optimality. As for data oriented mechanism, we have theoretically proved that it 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] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

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

[11] 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.