Data-Driven Pricing for Sensing Effort Elicitation in Mobile Crowd Sensing Systems

Haiming Jin, *Member, IEEE*, Baoxiang He, Lu Su, *Member, IEEE, ACM*, Klara Nahrstedt, *Fellow, IEEE, ACM*, and Xinbing Wang, *Senior Member, IEEE*

***Abstract*— The recent proliferation of human-carried mobile devices has given rise to mobile crowd sensing (MCS) systems that outsource sensory data collection to the public crowd. In order to identify truthful values from (crowd) workers' noisy or even conflicting sensory data, *truth discovery algorithms*, which jointly estimate workers' data quality and the underlying truths through quality-aware data aggregation, have drawn significant attention. However, the power of these algorithms could not be fully unleashed in MCS systems, unless workers' *strategic reduction of their sensing effort* is properly tackled. To address this issue, in this paper, we propose a *payment mechanism*, named Theseus, that deals with workers' such strategic behavior, and incentivizes high-effort sensing from workers. We ensure that, at the *Bayesian Nash Equilibrium* of the *non-cooperative game* induced by Theseus, all participating workers will spend their *maximum possible effort* on sensing, which improves their data quality. As a result, the aggregated results calculated subsequently by truth discovery algorithms based on workers' data will be highly accurate. Additionally, Theseus bears other desirable properties, including *individual rationality* and *budget feasibility*. We validate the desirable properties of Theseus through theoretical analysis, as well as extensive simulations.**

***Index Terms*— Incentive mechanism, mobile crowd sensing, truth discovery, sensing effort elicitation.**

I. Introduction

THE recent proliferation of increasingly capable human-carried mobile devices (e.g., smartphones, smartwatches) equipped with a plethora of on-board sensors (e.g., accelerometer, compass, GPS) has given rise to mobile crowd sensing (MCS). Different from traditional sensor networks [2], MCS is a new sensing paradigm which outsources sensory data collection to a crowd of participants, namely (crowd) workers. Thus far, a wide spectrum of MCS systems [3]–[5] have been deployed which cover almost every aspect of our lives, including smart transportation, healthcare, environmental monitoring, indoor localization, and many others.

In real practice, workers' sensory data are usually unreliable because of various factors (e.g., lack of effort, insufficient skill, poor sensor quality, background noise). Thus, the crowd sensing platform, which is usually a cloud-based central server, has to properly aggregate workers' noisy or even conflicting data so as to obtain accurate aggregated results. Clearly, a *weighted aggregation method* that assigns higher weights to workers with more reliable data is much more favorable than naive methods (e.g., averaging and voting) that view each worker equally, in that it shifts the aggregated results towards the data provided by more reliable workers.

The challenge, however, is that workers' reliability is usually unknown *a priori* by the platform, and should be inferred from the sensory data submitted by individual workers. To address this issue, *truth discovery*, which refers to a family of algorithms [6]–[9] that aim to discover meaningful facts from unreliable data, has been proposed and widely studied. Without any prior knowledge about workers' reliability, a truth discovery algorithm calculates jointly workers' weights and the aggregated results, based on the principles that the workers whose data are closer to the aggregated results will be assigned higher weights, and the data from a worker with a higher weight will be counted more in the aggregation.

However, *the power of truth discovery algorithms could not be fully unleashed in MCS systems, unless the platform properly deals with workers' strategic reduction of sensing effort*. This paper is exactly motivated by such fact due to the following reasons. Though yielding reasonably good performance under certain circumstances, truth discovery algorithms still suffer from the limitation that the aggregation accuracy highly depends on the quality of input data. If a vast majority of the data sources are unreliable, it will be hard or even impossible for these algorithms to obtain accurate aggregated results. This is exactly why past literature on truth discovery [6]–[9] assumes that most data sources have fairly good reliability. However, in MCS systems, such assumption does not hold, as the data sources here are *selfish* workers, who may *strategically reduce their costly sensing effort*, such as the time, resources, attention, and carefulness they put into the sensing tasks. Clearly, the level of a worker's sensing effort is among the major factors that affect her data quality. The reduction of workers' effort inevitably deteriorates

Manuscript received December 24, 2018; revised March 23, 2019 and July 15, 2019; accepted August 5, 2019; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor C. W. Tan. Date of publication November 8, 2019; date of current version December 17, 2019. This work was supported in part by National Key R&D Program of China under Grant 2017YFB1003000, Grant 2018YFB1004705, and Grant 2018YFB2100302, in part by the National Natural Science Foundation of China under Grant 61532012, Grant 61829201, and Grant 61902244, in part by the Science and Technology Innovation Program of Shanghai under Grant 17511105103 and Grant 18510761200, in part by the Shanghai Key Laboratory of Scalable Computing and Systems, in part by the Shanghai Municipal Science and Technology Commission under Grant 19YF1424600, and in part by the National Science Foundation under Grant CNS-1330491 and Grant 1652503. This work extends our article [1] published in ACM MobiHoc 2017. *(Corresponding author: Haiming Jin.)*

H. Jin is with the John Hopcroft Center for Computer Science, Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jinhaiming@sjtu.edu.cn).

B. He is with the UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: 1172548623@sjtu.edu.cn).

L. Su is with the Department of Computer Science and Engineering, The State University of New York at Buffalo, Buffalo, NY 14260 USA (e-mail: lusu@buffalo.edu).

K. Nahrstedt is with the Coordinated Science Laboratory, Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61820 USA (e-mail: klara@illinois.edu).

X. Wang is with the Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xwang8@sjtu.edu.cn).

Digital Object Identifier 10.1109/TNET.2019.2938453

the quality of their sensory data, which further impairs the aggregation accuracy. For example, in air quality monitoring applications [4], in order to save effort, workers may carry their mobile devices in their pockets instead of holding them on their hands as required, which may significantly degrade the reliability of their air quality measurements.

To address the aforementioned issue, in this paper, we take into consideration workers' strategic behavior, and propose a *payment mechanism*, named Theseus,[1] that offers payments to incentivize high-effort sensing from workers. Our workflow of an MCS system starts with the platform announcing the Theseus payment mechanism to workers before all the sensing happens. The workers' strategic behavior after the announcement of Theseus is then modeled using *game-theoretic methods*. In our model, Theseus induces a *non-cooperative game*,[2] called *sensing game*, where workers are the players who strategically decide their levels of effort for sensing. In order to elicit effort from workers, Theseus is then designed such that at the *Bayesian Nash Equilibrium (BNE)* of the sensing game, each participating worker maximizes her expected utility only when she spends her maximum possible effort. Clearly, Theseus improves the quality of workers' data by controlling a critical factor, that is, the level of their sensing effort. As a result, the aggregated results calculated subsequently by truth discovery algorithms based on workers' sensory data will be of high accuracy.

In summary, this paper makes the following contributions.

- In this paper, we propose a *payment mechanism*, called Theseus, which is used *in pair with a truth discovery algorithm* to ensure *high aggregation accuracy* in MCS systems with *continuous* and *categorical* sensing tasks by *dealing with workers' strategic reduction of sensing effort*.
- We derive *a suite of detailed parameter selection rules for Theseus* in both the *complete* and *incomplete information scenarios*, under which it could ensure *the existence of BNE in the sensing game*, *budget feasibility*, *individual rationality*, and *a small approximation ratio with high probability*.
- We derive the *BNEs of the non-cooperative sensing games induced by Theseus*, and prove that Theseus incentivizes workers to *spend their maximum possible sensing effort* at the derived BNEs.
- We demonstrate using *extensive numerical evaluations* that the proposed Theseus payment mechanism *ensures high aggregation accuracy* under a variety of parameter settings.

II. Related Work

In order to identify truthful values from workers' noisy or even conflicting sensory data in MCS systems, truth discovery algorithms [6]–[9], which jointly estimate workers' data quality and the underlying truths through quality-aware data aggregation, have drawn significant attention. However, *these algorithms usually cannot deal with workers' strategic reduction of sensing effort*, and thus, may yield unsatisfactory aggregation accuracy.

[1]The name Theseus comes from incenTivizing truth discovery with strategic data sources.

[2]Non-cooperative game refers to the family of games, where each player acts independently without collaboration or communication with others, whereas, in cooperative games, players may communicate with each other and form coalitions.

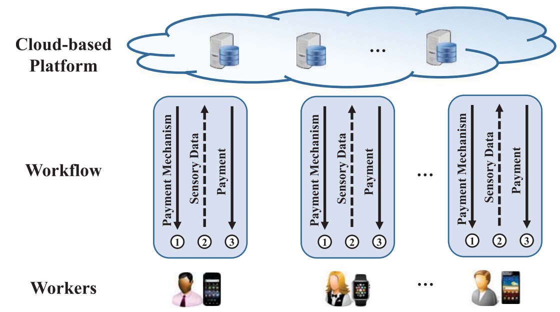

Fig. 1. Interaction between the platform and workers (where circled numbers represent the order of the events).

Another line of prior work related to this paper is a series of incentive mechanisms [10]–[37] recently developed by the research community in order to stimulate worker participation in MCS systems. Most of these past literature [10]–[33] adopts game-theoretic methods [38]–[40], due to their ability to deal with workers' strategic behavior. Among them, auction-based incentive mechanisms [16]–[29] typically consider workers' strategic bidding of the prices and sensing task choices to the platform. Furthermore, some prior work [30]–[32] tackles workers' strategic manipulation of reported private and sensitive data due to privacy concerns. However, *none of them study workers' strategic reduction of sensing effort as in this work*. Mechanisms that elicit effort from crowd workers have been investigated in past literature [11]–[16], but *none of them is designed to work in pair with truth discovery algorithms*. Note that although there exist incentive mechanisms [35]–[37] that work jointly with truth discovery algorithms, different from this paper, [35] is not based on game-theoretic models, and thus, cannot tackle workers' strategic behavior, [37] aims to design sybil-resistant truth discovery algorithms, and [36] specifically investigates MCS systems with copiers.

Different from prior work, we design a payment mechanism, which is used *in pair with a truth discovery algorithm* to ensure *high aggregation accuracy* by *incentivizing workers to spend their maximum possible sensing effort*.

III. Preliminaries

In this section, we introduce the system overview, truth discovery algorithms, our game theoretic model, as well as the design objectives.

A. System Overview

We consider an MCS system consisting of a cloud-based platform, and a set of S potential participating workers, denoted as $\mathcal{S} = \{1, 2 \cdots, S\}$. The platform holds a set of M sensing tasks, denoted as $\mathcal{M} = \{1, 2, \cdots, M\}$, and each task requires workers to sense a particular object, event, or phenomenon locally, and report to the platform the sensory data in the form of *continuous values*. Such MCS systems collecting continuous data from the crowd, constitute a large portion of the currently deployed MCS systems, such as environmental monitoring applications that collect air quality or noise level measurements from participating workers. The scenario where workers report *categorical sensory data* to the platform will be investigated in Section VII. We demonstrate the interaction between the platform and workers in Figure 1, and describe the complete workflow of our MCS system model as follows.

- Firstly, the platform announces the set of sensing tasks $\mathcal{M}$, as well as the payment mechanism, to the set of all potential participating workers $\mathcal{S}$ (step ①).
- After such announcements, each worker $s \in \mathcal{S}$ decides whether or not to participate in the sensing tasks. Then, the workers that choose to participate decide the levels of their sensing effort (e.g., time, resources, attention, carefulness), and carry out sensing according to the decided effort levels. We denote the set of participating workers as $\mathcal{S}' \subseteq \mathcal{S}$. Each worker $s \in \mathcal{S}'$ then submits to the platform the sensory data x_m^s for each task $m \in \mathcal{M}$ upon completion of sensing[3] (step ②).
- After receiving workers' data, the platform pays each participating worker according to the payment calculated using the payment mechanism (step ③).
- Finally, based on the collected data, the platform calculates an aggregated result x_m^* for each task m, and uses it as an estimate for the ground truth x_m^{truth}, which is unknown to both the platform and the workers.

As the quality of different workers' sensory data typically varies, an ideal approach is to use a weighted aggregation scheme which assigns higher weights to workers with higher data quality. However, in practice, workers' data quality is usually unknown *a priori* to the platform. Therefore, in our model, the platform utilizes one of the *truth discovery* algorithms [6]–[9] to aggregate workers' data, which calculates workers' weights and estimates the ground truths in a joint manner. An introduction of such algorithms is provided in the following Section III-B.

B. Truth Discovery

Although existing truth discovery algorithms [6]–[9] differ in their specific ways to calculate workers' weights and the aggregated results, their common procedure could be summarized as in the following Algorithm 1.

A truth discovery algorithm, as described in Algorithm 1, typically starts with a random guess of tasks' ground truths, and then iteratively updates workers' weights, as well as the estimated ground truths until convergence.

Weight Calculation: In this step, tasks' estimated ground truths are assumed to be fixed, and the weight w_s of each worker $s \in \mathcal{S}'$ is calculated as

$$w_s = \omega\left(\sum_{m\in\mathcal{M}} d(x_m^s, x_m^*)\right), \tag{1}$$

where $\omega(\cdot)$ is some monotonically decreasing function, and $d(\cdot)$ denotes the function that calculates the distance between the worker's data x_m^s and the estimated ground truth x_m^*. Although different truth discovery algorithms may adopt different functions $\omega(\cdot)$ and $d(\cdot)$, they share the same underlying principle that higher weights are assigned to workers whose data are closer to the estimated ground truths.

[3]Clearly, in practice, each individual worker may not be able to execute all the sensing tasks hosted by the platform. Thus, a more realistic model is to introduce an affinity term for each worker-task pair (s, m) that indicates whether or not worker s is able to execute task m. However, to simplify the presentation of our subsequent mathematical analyses, we assume that each worker is capable to execute all the tasks.

Algorithm 1 Truth Discovery Algorithm

```
Input: Workers' data {x_m^s | m ∈ M, s ∈ S'};
Output: Estimated ground truths {x_m^* | m ∈ M};
Randomly initialize the ground truth for each task;
repeat
    // Weight calculation
    foreach s ∈ S' do
        Update the weight w_s based on current estimated
        ground truths using Equation (1);
    // Truth estimation
    foreach m ∈ M do
        Update the estimated ground truth x_m^* based on
        workers' current weights using Equation (2);
until Convergence criterion is satisfied;
return Estimated ground truths {x_m^* | m ∈ M};
```

Truth Estimation: In this step, workers' weights are assumed to be fixed, and the estimated ground truth x_m^* of each task m is derived as

$$x_m^* = \frac{\sum_{s\in\mathcal{S}'} w_s x_m^s}{\sum_{s\in\mathcal{S}'} w_s}. \tag{2}$$

In such weighted aggregation method, the aggregated result x_m^* relies more on the workers with higher weights. Usually, the convergence criterion is application specific. For example, the algorithm could be treated as converged as long as the difference between the estimated ground truths in two consecutive iterations is less than a threshold.

Note that the payment mechanism that we propose in this paper is independent with the specific forms of the functions $\omega(\cdot)$ and $d(\cdot)$ in Equation (1). Therefore, it is able to work jointly with any truth discovery algorithm that shares the same procedure as Algorithm 1. Further discussions on this point will be provided in Section IV.

C. Game Theoretic Model

As the aggregation accuracy of truth discovery algorithms highly depends on the quality of input data, existing work on truth discovery [6]–[9] assumes that most data sources have fairly good reliability. In MCS systems, however, such assumption does not hold, as the data sources here are usually *strategic* and *selfish* workers, who may reduce their sensing effort strategically, and thus, provide unreliable data.

In this paper, we take into consideration workers' strategic behavior, and incentivize workers to provide high quality data using a *payment mechanism* defined in Definition 1.

Definition 1 (Payment Mechanism): A payment mechanism, denoted as $p : \mathcal{X} \to \mathbb{R}^S$, where $\mathcal{X}$ denotes the set containing all possible sets of workers' sensory data, calculates the payments to workers based on the collected set of data $\mathbf{x} = \{x_m^s | m \in \mathcal{M}, s \in \mathcal{S}'\}$. We use $p_s(\mathbf{x}) \geq 0$ to denote the payment to worker s, when the set of collected data is $\mathbf{x}$. Note that $p_s(\mathbf{x}) = 0$, if worker s drops out.

A payment mechanism defined in Definition 1 is data-driven in the sense that it calculates the payments to workers based on their sensory data. As mentioned in Section III-A, the platform firstly announces to workers the payment mechanism $p(\cdot)$, which then induces a *non-cooperative game*,[4] referred to as *sensing game* in the rest of this paper, where workers are the players. In this game, each worker decides whether or not to participate by evaluating her own expected utility. That is, a worker s will drop out, if participation leads to a negative expected utility, and otherwise, she will participate with a specific effort level e_s that maximizes her expected utility. Similar to past literature [6], [41], we assume that the difference between any worker s's data and the ground truth follows a zero-mean Gaussian distribution, i.e.,

$$X_m^s - X_m^{\text{truth}} \sim N(0, \delta_s^2), \tag{3}$$

where X_m^s and X_m^{truth} are the random variables corresponding to x_m^s and the ground truth x_m^{truth} respectively, and $N(0, \delta_s^2)$ denotes a Gaussian random variable with mean zero and standard deviation δ_s. Although we assume that such difference follows a Gaussian distribution, the results in this paper could be generalized, with some adaptation, to scenarios with other types of distributions. Clearly, the standard deviation δ_s captures a worker s's data quality, as the less the value of δ_s, the more likely that her sensory data will be close to the ground truth.

As a worker's data quality typically increases with her effort level, we assume that $\delta_s = q_s(e_s) \in [\underline{\delta}_s, \overline{\delta}_s]$ for each worker s, where $q_s(\cdot)$ is a bounded monotonically non-increasing function. We allow, in our model, workers to have different $q_s(\cdot)$ functions and ranges for their δ_s's, because apart from a worker's effort, her data quality is also affected by other factors (e.g., skill level, sensor quality, environment noise). As each worker s is assigned a single weight w_s in the truth discovery algorithm adopted by us (Algorithm 1), we assume that she spends the same amount of effort e_s on all the tasks. We leave the study of the scenario where workers have different effort levels on different tasks in our future work.

For simplicity, we use δ_s instead of e_s as a worker s's *strategy*, and use $\delta_s = \perp$ to denote that the worker chooses to drop out. Thus, a worker s's *strategy space* is $[\underline{\delta}_s, \overline{\delta}_s] \cup \{\perp\}$. As given by Equation (3), the distribution of any worker s's data depends on δ_s, we use $\mathbf{x}(\boldsymbol{\delta})$ to denote the set of collected data, and $\mathbf{X}(\boldsymbol{\delta})$ the random variable corresponding to $\mathbf{x}(\boldsymbol{\delta})$, when workers' *strategy profile* is $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_S)$. Then, we define a worker's utility in Definition 2.

Definition 2 (Worker's Utility): Given the payment mechanism $p(\cdot)$ and workers' strategy profile $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_S)$, any worker s's utility is

$$u_s(\boldsymbol{\delta}) = p_s\big(\mathbf{x}(\boldsymbol{\delta})\big) - C_s(\delta_s), \tag{4}$$

where $C_s(\cdot)$ is a monotonically decreasing function for $\delta_s \in [\underline{\delta}_s, \overline{\delta}_s]$, and $C_s(\perp) = 0$. $C_s(\delta_s)$ denotes worker s's sensing cost when her strategy is δ_s. Therefore, the expected utility of worker s (evaluated by worker s) is

$$\mathbb{E}_{\boldsymbol{\delta}_{-s}}\big[u_s(\delta_s, \boldsymbol{\delta}_{-s})\big] = \mathbb{E}_{\boldsymbol{\delta}_{-s}}\big[p_s\big(\mathbf{X}(\delta_s, \boldsymbol{\delta}_{-s})\big)\big] - C_s(\delta_s), \tag{5}$$

where $\boldsymbol{\delta}_{-s} = (\delta_1, \cdots, \delta_{s-1}, \delta_{s+1}, \cdots, \delta_S)$ denotes workers' strategy profile excluding δ_s.

In general cases, the calculation of a worker s's expected utility in Equation (5) requires the knowledge of the joint distribution of $\boldsymbol{\delta}_{-s}$. However, because of the specific design of our payment mechanism described in Section V, the calculation can be done without knowing such joint distribution. We leave the detailed discussion on the required prior statistical knowledge in Section VI.

[4]Refer to Footnote 2 for definition.

D. Design Objectives

In this paper, we aim to design a payment mechanism which preserves several desirable properties at the *Bayesian Nash Equilibrium (BNE)*, formally defined in Definition 3, of the sensing game.

Definition 3 (BNE): The strategy profile $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*, \cdots, \delta_S^*)$ is a Bayesian Nash Equilibrium (BNE) of the sensing game, if

$$\mathbb{E}_{\boldsymbol{\delta}^*_{-s}}\big[u_s(\delta_s^*, \boldsymbol{\delta}^*_{-s})\big] \geq \mathbb{E}_{\boldsymbol{\delta}^*_{-s}}\big[u_s(\delta_s, \boldsymbol{\delta}^*_{-s})\big], \quad \forall s \in \mathcal{S}, \delta_s, \tag{6}$$

where $\boldsymbol{\delta}^*_{-s} = (\delta_1^*, \cdots, \delta_{s-1}^*, \delta_{s+1}^*, \cdots, \delta_S^*)$.

Clearly, BNE $\boldsymbol{\delta}^*$ satisfies that any worker s maximizes her expected utility by taking strategy δ_s^* given that other workers take strategies $\boldsymbol{\delta}^*_{-s}$. One desirable property we aim to achieve is *individual rationality* defined in Definition 4.

Definition 4 (Individual Rationality): A payment mechanism $p(\cdot)$ is individual rational, if and only if no worker has negative expected utility at BNE $\boldsymbol{\delta}^*$, i.e.,

$$\mathbb{E}_{\boldsymbol{\delta}^*_{-s}}\big[u_s(\delta_s^*, \boldsymbol{\delta}^*_{-s})\big] \geq 0, \quad \forall s \in \mathcal{S}. \tag{7}$$

The property of individual rationality is necessary for a payment mechanism, as it prevents workers from being disincentivized to participate. Because usually, in practice, the platform works under a fixed budget, another design objective considered is *budget feasibility* defined in Definition 5.

Definition 5 (Budget Feasibility): A payment mechanism $p(\cdot)$ is budget feasible, if and only if the expected overall payment at BNE $\boldsymbol{\delta}^*$ does not exceed the budget B, i.e.,

$$\mathbb{E}_{\boldsymbol{\delta}^*}\Bigg[\sum_{s \in \mathcal{S}'} p_s\big(\mathbf{X}(\boldsymbol{\delta}^*)\big)\Bigg] \leq B. \tag{8}$$

Another critical desirable property is that workers at BNE provide high quality data, so that the truth discovery algorithm ensures low *error probability*, defined in Definition 6.

Definition 6 (Error Probability): Given any $\alpha > 0$, we define the error probability of a truth discovery algorithm as

$$\Pr\Bigg(\frac{1}{M}\sum_{m=1}^{M}\big|X_m^* - X_m^{\text{truth}}\big| \geq \alpha\Bigg), \tag{9}$$

where X_m^* denotes the random variable corresponding to the estimated ground truth x_m^*. Clearly, it is the probability that the mean absolute error (MAE), $\frac{1}{M}\sum_{m=1}^{M}\big|X_m^* - X_m^{\text{truth}}\big|$, of a truth discovery algorithm is no less than a threshold α.

In summary, our objective is to design an *individual rational* and *budget feasible payment mechanism*, which ensures that the truth discovery algorithm guarantees *low error probability at BNE*.

Table I summarizes the notations frequently used in this paper, where some will be introduced later.

IV. Mathematical Formulation

In this section, we formally formulate the payment mechanism design problem mathematically. Firstly, we introduce the following Lemma 1 that establishes an upper bound for the error probability of a truth discovery algorithm defined in Definition 6.

TABLE I
NOTATIONS

Notation	Definition
$\mathcal{S}$, S	the set and number of workers
$\mathcal{S}'$	the set of participating workers
$\mathcal{M}$, M	the set and number of tasks
$\mathbf{x}$	the collected set of data
$p(\cdot)$	payment mechanism
$\boldsymbol{\delta}$, $\boldsymbol{\delta}^*$	strategy profile and BNE for continuous MCS tasks
$\mathbf{e}$, $\mathbf{e}^*$	strategy profile and BNE for categorical MCS tasks
$u_s(\cdot)$	worker s's utility function

Lemma 1: Given any $\alpha > 0$ and workers' strategy profile $\boldsymbol{\delta} = (\delta_1, \delta_2, \cdots, \delta_S)$, we have that

$$\Pr\left(\frac{1}{M}\sum_{m=1}^{M}\left|X_m^* - X_m^{\text{truth}}\right| \geq \alpha\right) \leq \sqrt{\frac{2}{\pi}}\frac{\sum_{s\in\mathcal{S}'}\delta_s}{\alpha}, \quad (10)$$

that is, the error probability of a truth discovery algorithm is upper bounded by $\sqrt{\frac{2}{\pi}}\frac{\sum_{s\in\mathcal{S}'}\delta_s}{\alpha}$.

Proof: Please refer Appendix A for the detailed proof. □

Given any fixed α, the upper bound of the error probability of a truth discovery algorithm given by Lemma 1 is proportional to $\sum_{s\in\mathcal{S}'}\delta_s$, i.e., the sum of all participating workers' δ_s's. Thus, we aim to minimize $\sum_{s\in\mathcal{S}'}\delta_s^*$ in order to get a good guarantee for the error probability at BNE $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*, \cdots, \delta_S^*)$. The formal mathematical formulation of the *payment mechanism design (PMD)* problem is given in the following optimization program.

PMD Problem:

$$\min_{p(\cdot)\in\mathcal{P}} \sum_{s\in\mathcal{S}'}\delta_s^* \quad (11)$$

$$\text{s.t. } \mathbb{E}_{\boldsymbol{\delta}^*_{-s}}\left[u_s(\delta_s^*, \boldsymbol{\delta}^*_{-s})\right] \geq 0, \quad \forall s\in\mathcal{S} \quad (12)$$

$$\mathbb{E}_{\boldsymbol{\delta}^*}\left[\sum_{s\in\mathcal{S}'} p_s\big(\mathbf{X}(\boldsymbol{\delta}^*)\big)\right] \leq B \quad (13)$$

Constants: The PMD problem takes as inputs the worker set $\mathcal{S}$, the budget B, as well as the set $\mathcal{P}$, which denotes the set consisting of all the possible payment mechanisms, such that a BNE exists for the corresponding sensing game.

Variable: The variable of the PMD problem is the payment mechanism $p(\cdot)$. Furthermore, $\boldsymbol{\delta}^*$ denotes the BNE corresponding to $p(\cdot)$, and $\mathcal{S}'$ and $\mathbf{X}(\boldsymbol{\delta}^*)$ denote, respectively, the set of participating workers and collected sensory data at the BNE $\boldsymbol{\delta}^*$. Note that as $\boldsymbol{\delta}^*$, δ_s^*, $\boldsymbol{\delta}^*_{-s}$, and $\mathcal{S}'$ in the PMD problem are determined by $p(\cdot)$, more comprehensive notations of them are $\boldsymbol{\delta}^*(p(\cdot))$, $\delta_s^*(p(\cdot))$, $\boldsymbol{\delta}^*_{-s}(p(\cdot))$, and $\mathcal{S}'(p(\cdot))$, respectively. For simplicity, however, we denote them as $\boldsymbol{\delta}^*$, δ_s^*, $\boldsymbol{\delta}^*_{-s}$, and $\mathcal{S}'$ as in the PMD problem.

Objective Function: The objective (Equation (11)) of the PMD problem is to find the payment mechanism from $\mathcal{P}$ with the minimum $\sum_{s\in\mathcal{S}'}\delta_s^*$ at the corresponding BNE $\boldsymbol{\delta}^*$, which is equivalent to minimizing the upper bound, as derived in Lemma 1, of a truth discovery algorithm's error probability at BNE for a fixed α.

Constraints: Constraint (12) and (13) ensure, respectively, that any feasible solution $p(\cdot)$ to the PMD problem satisfies individual rationality and budget feasibility.

Thus, the PMD problem aims to find the individual rational and budget feasible payment mechanism, which minimizes the upper bound (given by Lemma 1) of a truth discovery algorithm's error probability at the corresponding BNE for any fixed α. Clearly, our formulation of the PMD problem is valid for an arbitrary way of assigning workers' weights. Therefore, the above formulation and the proposed payment mechanism to be presented in the following section can be applied to any truth discovery algorithm that has the same procedure as Algorithm 1.

V. PROPOSED PAYMENT MECHANISM

As solving directly the optimal payment mechanism is hard, in this section, we propose our own payment mechanism, named Theseus, in Algorithm 2, which approximately solves the PMD problem with good performance guarantees.

Algorithm 2 Theseus Payment Mechanism

```
Input: M, S, S', x, {(a_s, b_s) | s ∈ S};
Output: {p_s | s ∈ S};
1 foreach worker s ∈ S do
2     if s ∈ S' then
3         Randomly pick another worker r ∈ S';
4         p_s ← b_s − a_s (1/M) Σ_{m=1}^{M} (x_m^s − x_m^r)^2;
5     else
6         p_s ← 0;
7 return {p_s | s ∈ S};
```

Algorithm 2 takes as inputs the set of tasks $\mathcal{M}$, workers $\mathcal{S}$, and participating workers $\mathcal{S}'$, as well as the set of collected sensory data $\mathbf{x}$, and $\{(a_s, b_s)|s \in \mathcal{S}\}$ where a_s and b_s are positive parameters related to the payment to worker s. The calculation of the payment to any participating worker (line 2-4) borrows the high-level idea of the peer prediction method [42], which basically decides the payment based on the difference between her data and that of a randomly selected *reference worker*. That is, if worker s participates (i.e., $s\in\mathcal{S}'$), Algorithm 2 randomly picks another reference worker r from the set of participating workers $\mathcal{S}'$ (line 3). Next, the payment p_s to this worker s is set as

$$p_s = b_s - a_s\frac{1}{M}\sum_{m=1}^{M}(x_m^s - x_m^r)^2. \quad (14)$$

Clearly, the more worker s's data agrees with that of the randomly selected reference worker r, the higher her payment p_s will be. If any worker s drops out (i.e., $s \notin \mathcal{S}'$), the algorithm will set her payment as 0 (line 6). Finally, the algorithm returns the set of payments to all workers $\{p_s|s\in\mathcal{S}\}$ (line 7). By now, our description of Theseus has been finished except for one missing piece, that is, how the parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ are set, which is presented in the following Section VI.

Clearly, another intuitive way of deciding the payment p_s to each participating worker s is to set p_s to be positively correlated to her weight w_s calculated by the truth discovery algorithm using Equation (1). However, we do not adopt this approach due to the difficulty in analyzing the properties of the induced sensing game.

VI. Parameterization

In this section, we introduce our careful selection of the parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ in order to ensure that Theseus achieves good performance. To simplify our analysis, we assume that each worker s's cost function $C_s(\cdot)$ is linear in $\delta_s \in [\underline{\delta}_s, \overline{\delta}_s]$, i.e.,

$$C_s(\delta_s) = -c_{s,1}\delta_s + c_{s,2}, \quad \forall \delta_s \in [\underline{\delta}_s, \overline{\delta}_s], \tag{15}$$

where $c_{s,1}$ and $c_{s,2}$ are positive parameters. Note that such selection of each worker s's cost function conforms to the requirement that her cost should decrease with the increase of δ_s.

According to how much prior knowledge the platform has about workers' cost functions, we parameterize Theseus in the following two scenarios, namely the *complete information scenario* where the platform knows exactly each worker s's $c_{s,1}$ and $c_{s,2}$ (Section VI-A), as well as the *incomplete information scenario* where only limited information about $c_{s,1}$ and $c_{s,2}$ is known by the platform (Section VI-B). In both scenarios, we assume that $\underline{\delta}_1, \underline{\delta}_2, \cdots, \underline{\delta}_S$, i.e, the lower bounds of workers' δ_s's, are i.i.d. random variables within the range $[\underline{\delta}, \overline{\delta}]$ with PDF $f(\cdot)$. Furthermore, the PDF $f(\cdot)$ is assumed to be *a priori* known by the platform and workers, which, as will be shown in Section VI-A2 and VI-B2, is the only prior statistical knowledge needed to evaluate workers' expected utilities.

A. Complete Information Scenario

1) Parameter Selection: As aforementioned, in this section, we assume that the platform knows exactly both $c_{s,1}$ and $c_{s,2}$ in each worker s's cost function. Although, in practice, it might be hard for the platform to obtain such exact knowledge, the complete information scenario is still relevant and interesting to study, because it sheds light upon the philosophy of parameterizing Theseus in the incomplete information scenario in Section VI-B. For any given $\Delta_t \in [\underline{\delta}, \overline{\delta}]$, we can parameterize Theseus with any set of parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ that satisfy Condition (16)-(18).

$$\begin{cases} a_s \geq \dfrac{c_{s,1}}{2\underline{\delta}}, & \forall s \in \mathcal{S} \quad (16) \\ b_s = a_s\big(\Delta_t^2 + A(\Delta_t)\big) - c_{s,1}\Delta_t + c_{s,2}, & \forall s \in \mathcal{S} \quad (17) \\ \displaystyle\sum_{s=1}^{S} b_s \leq B + \sum_{s=1}^{S} 2a_s\underline{\delta}^2, & (18) \end{cases}$$

where $A(\Delta_t) = \int_{\underline{\delta}}^{\Delta_t} u^2 \frac{f(u)}{\int_{\underline{\delta}}^{\Delta_t} f(v)dv} du$. The criterion of selecting the additional parameter Δ_t will be discussed in Section VI-A2 as we analyze the performance guarantees of the parameter selection given by Condition (16)-(18). For each $s \in \mathcal{S}$, as b_s is exactly determined by a_s due to Condition (17), one way of parameter selection is to choose an $a_s \geq \frac{c_{s,1}}{2\underline{\delta}}$ such that Condition (18) is satisfied.

2) Analysis: In this section, we carry out analyses about the desirable properties of Theseus by parameterizing it according to Condition (16)-(18). We derive the BNE of the sensing game that corresponds to such parameterization in the following Theorem 1.

Theorem 1: If parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ satisfy Condition (16) and (17), we have that $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*, \cdots, \delta_S^*)$, where, for each worker $s \in \mathcal{S}$,

$$\delta_s^* = \begin{cases} \perp, & \text{if } \underline{\delta}_s > \Delta_t \\ \underline{\delta}_s, & \text{if } \underline{\delta}_s \leq \Delta_t, \end{cases} \tag{19}$$

is a BNE of the sensing game in the complete information scenario.

Proof: Please refer to Appendix B for detailed proof. □

Theorem 1 gives us a BNE of the sensing game, where every worker s with $\underline{\delta}_s > \Delta_t$ will voluntarily drop out, and as long as $\underline{\delta}_s \leq \Delta_t$, the worker s will participate with strategy $\underline{\delta}_s$, which is exactly the smallest standard deviation of the difference between her data and the ground truths. That is, by satisfying Condition (16) and (17), Theseus will only incentivize workers who potentially is capable of providing high quality data to participate, and those who choose to participate will exert their maximum amount of effort, leading them to provide reliable data. Note that there might be multiple BNEs for the sensing game. However, to the best of our knowledge, we have not found other BNEs except for the one given in Theorem 1, on which our further analyses in this section are based. We leave the derivation of other BNEs or the proof of the uniqueness of BNE in our future work. Next, we prove in the following Theorem 2 that Theseus satisfies budget feasibility in the complete information scenario by satisfying Condition (18).

Theorem 2: Condition (18) ensures that Theseus is budget feasible in the complete information scenario.

Proof: Please refer to the supplementary material [43] for the detailed proof. □

Clearly, as stated in the following Theorem 3, Theseus satisfies individual rationality in the complete information scenario.

Theorem 3: Theseus is individual rational in the complete information scenario.

Proof: Please refer to the supplementary material [43] for the detailed proof. □

Next, we discuss our selection criterion of the parameter Δ_t. Following notational conventions in order statistics, we denote $\underline{\delta}_{(1)} = \min\{\underline{\delta}_1, \underline{\delta}_2, \cdots, \underline{\delta}_S\}$. We assume that the CDF $F(\cdot)$ of any $\underline{\delta}_s$ is invertible, and its inverse is $F^{-1}(\cdot)$. Based on Theorem 1, if Δ_t is set to be too small, no workers will participate at the BNE. Thus, we establish a lower bound for Δ_t in the following Theorem 4.

Theorem 4: Given any $\theta_c \in (0, 1)$, if $\Delta_t \geq F^{-1}\big(1 - \sqrt[S]{1 - \theta_c}\big)$, then $\Pr\big(\underline{\delta}_{(1)} \leq \Delta_t\big) \geq \theta_c$, i.e., the probability that at least one worker chooses to participate at the BNE of the sensing game, in the complete information scenario, is no less than the threshold θ_c.

Proof: Please refer to Appendix C for the detailed proof. □

In the rest of our analyses, we use APP to denote the value of the PMD problem's objective function guaranteed by Theseus. Theorem 4 gives us that if $\Delta_t \geq F^{-1}\big(1 - \sqrt[S]{1 - \theta_c}\big)$, the probability that there exists at least one participating worker at the BNE of the sensing game is guaranteed to be no less than the predefined threshold $\theta_c \in (0, 1)$. However, this does not mean that Δ_t could be infinitely large, because the greater Δ_t is, the farther APP will drift apart from the minimum value of the PMD problem's objective function. Thus, in the following Theorem 5, we derive an upper bound for the parameter Δ_t. Note that for any payment mechanism that ensures the participation of at least one worker, the minimum possible value for the objective function is $\text{OPT} = \underline{\delta}_{(1)}$, which is the optimal benchmark that we compare APP with.

Theorem 5: In the complete information scenario, given $\alpha_c > 1$ and $\beta_c \in (0,1)$, we have that

$$\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_c\right) \leq \beta_c, \tag{20}$$

if $\Delta_t \leq \overline{\Delta}_t$, where $\overline{\Delta}_t$ is the solution to

$$\overline{\Delta}_t + \sqrt{-\frac{2}{S \ln \beta_c}}\Big(R\big(\overline{\Delta}_t\big)S - \underline{\delta}\alpha_c\Big) = 0, \tag{21}$$

with $R\big(\overline{\Delta}_t\big) = \int_{\underline{\delta}}^{\overline{\Delta}_t} u \frac{f(u)}{\int_{\underline{\delta}}^{\overline{\Delta}_t} f(v)dv} du$.

Proof: Please refer to Appendix D for the detailed proof. □

By Theorem 5, we have that, as long as $\Delta_t \leq \overline{\Delta}_t$, the probability that the approximation ratio $\frac{\text{APP}}{\text{OPT}} \geq \alpha_c$ is no greater than β_c, for the predefined constants $\alpha_c > 1$ and $\beta_c \in (0,1)$. This shows the probabilistic guarantee on the approximation ratio of Theseus compared to the optimal payment mechanism. Next, we have the following Corollary 1 about the range from which the parameter Δ_t should be selected.

Corollary 1: By jointly considering Theorem 4 and 5, Δ_t should satisfy $F^{-1}\big(1 - \sqrt[S]{1-\theta_c}\big) \leq \Delta_t \leq \overline{\Delta}_t$ in the complete information scenario, so as to guarantee that with high probability there exist participating workers at the corresponding BNE (Theorem 4), and that with high probability Theseus has a small approximation ratio (Theorem 5).

B. Incomplete Information Scenario

1) Parameter Selection: In this section, we study a more practical incomplete information scenario, where the platform does not know the exact values of each worker s's $c_{s,1}$ and $c_{s,2}$, but instead, only knows that $c_{s,1} \in [\underline{c}_1, \overline{c}_1]$, and $c_{s,2} \in [\underline{c}_2, \overline{c}_2]$, for each worker s. In this case, given any Δ_l and Δ_h, such that $\underline{\delta} \leq \Delta_l < \Delta_h \leq \overline{\delta}$, we can parameterize Theseus with any set of parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ such that Condition (22)-(25) are satisfied.

$$\begin{cases} a_s \geq \dfrac{\overline{c}_1}{2\underline{\delta}}, & \forall s \in \mathcal{S} \quad (22) \\ b_s \leq a_s\big(\Delta_h^2 + A(\Delta_h)\big) - \overline{c}_1\Delta_h + \underline{c}_2, & \forall s \in \mathcal{S} \quad (23) \\ b_s \geq a_s\big(\Delta_l^2 + A(\Delta_h)\big) - \underline{c}_1\Delta_l + \overline{c}_2,, & \forall s \in \mathcal{S} \quad (24) \\ \displaystyle\sum_{s=1}^{S} b_s \leq B + \sum_{s=1}^{S} 2a_s\underline{\delta}^2, & (25) \end{cases}$$

where $A(\Delta_h) = \int_{\underline{\delta}}^{\Delta_h} u^2 \frac{f(u)}{\int_{\underline{\delta}}^{\Delta_h} f(v)dv} du$. Note that the criterion of selecting Δ_l and Δ_h will be discussed in Section VI-B2 as we introduce the corresponding analyses. Given these conditions, one specific way of parameter selection for each $s \in \mathcal{S}$ is to choose an $a_s \geq \frac{\overline{c}_1}{2\underline{\delta}}$ such that Condition (22)-(25) are satisfied.

2) Analysis: In this section, we firstly characterize the BNE of the sensing game by parameterizing Theseus in the incomplete information scenario according to Condition (22)-(25) in the following Theorem 6.

Theorem 6: If parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ satisfy Condition (22)-(24), we have a BNE $\boldsymbol{\delta}^* = (\delta_1^*, \delta_2^*, \cdots, \delta_S^*)$ of the sensing game in the incomplete information scenario, such that, for each worker $s \in \mathcal{S}$,

$$\delta_s^* = \begin{cases} \perp, & \text{if } \underline{\delta}_s > \Delta_h \\ \underline{\delta}_s, & \text{if } \underline{\delta}_s \leq \Delta_l. \end{cases} \tag{26}$$

Proof: Please refer to Appendix E for the detailed proof. □

Theorem 6 characterizes a BNE of the sensing game, where each worker s with $\underline{\delta}_s > \Delta_h$ will drop out, and as long as $\underline{\delta}_s \leq \Delta_l$, she will participate with strategy $\underline{\delta}_s$. Note that, at the BNE, each worker s with $\underline{\delta}_s \in (\Delta_l, \Delta_h]$ has to evaluate her expected utility based on the specific choice of $\{(a_s, b_s)|s \in \mathcal{S}\}$ in order to make the decision of whether or not to participate. All of the following analyses in this section are based on the BNE characterized in Theorem 6. We also leave the proof of the uniqueness of BNE or the derivation of other BNEs in our future work. Next, we introduce in Theorem 7 and 8 about the budget feasibility and individual rationality of Theseus in the incomplete information scenario.

Theorem 7: Condition (25) ensures that Theseus is budget feasible in the incomplete information scenario.

Theorem 8: Theseus is individual rational in the incomplete information scenario.

The proof of Theorem 7 is the same as that of Theorem 2 except that the Δ_t is replaced by Δ_h, and the proof of Theorem 8 is exactly identical to that of Theorem 3. Thus, we omit the formal proofs of Theorem 7 and 8 in this paper. Similar to Section VI-A2, we establish ranges from which we select parameters Δ_l and Δ_h. In the following Theorem 9, we introduce a lower bound for Δ_l.

Theorem 9: Given any $\theta_{ic} \in (0,1)$, if $\Delta_l \geq F^{-1}\big(1 - \sqrt[S]{1-\theta_{ic}}\big)$, then $\Pr\big(\underline{\delta}_{(1)} \leq \Delta_l\big) \geq \theta_{ic}$, i.e., the probability that at least one worker chooses to participate at the BNE of the sensing game, in the incomplete information scenario, is no less than the threshold θ_{ic}.

The proof of Theorem 9 is omitted in this paper as well, because it can be directly adapted from that of Theorem 4 by changing Δ_t to Δ_l and θ_c to θ_{ic}. By Theorem 9, we have that, in the incomplete information scenario, it is Δ_l that decides the probability that at least one worker chooses to participate at the BNE of the sensing game. That is, as long as $\Delta_l \geq F^{-1}\big(1 - \sqrt[S]{1-\theta_{ic}}\big)$, this probability, i.e., $\Pr\big(\underline{\delta}_{(1)} \leq \Delta_l\big)$, will be no less than the predefined threshold θ_{ic}. Next, in the following Theorem 10, where APP and OPT have the same meanings as in Theorem 5, we derive an upper bound for Δ_h.

Theorem 10: In the incomplete information scenario, given $\alpha_{ic} > 1$ and $\beta_{ic} \in (0,1)$, we have that

$$\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_{ic}\right) \leq \beta_{ic}, \tag{27}$$

if $\Delta_h \leq \overline{\Delta}_h$, where $\overline{\Delta}_h$ is the solution to

$$\overline{\Delta}_h + \sqrt{-\frac{2}{S \ln \beta_{ic}}}\Big(R\big(\overline{\Delta}_h\big)S - \underline{\delta}\alpha_{ic}\Big) = 0, \tag{28}$$

with $R\big(\overline{\Delta}_h\big) = \int_{\underline{\delta}}^{\overline{\Delta}_h} u \frac{f(u)}{\int_{\underline{\delta}}^{\overline{\Delta}_h} f(v)dv} du$.

Proof: Please refer to Appendix F for the detailed proof. □

Similar to Theorem 5, Theorem 10 gives us a probabilistic guarantee on the approximation ratio of Theseus compared to the optimal payment mechanism in the incomplete information scenario. That is, as long as $\Delta_h \leq \overline{\Delta}_h$, the probability that the approximation ratio $\frac{\text{APP}}{\text{OPT}} \geq \alpha_{ic}$ is no greater than β_{ic}, for the predefined constants $\alpha_{ic} > 1$ and $\beta_{ic} \in (0,1)$. Next, we introduce in Corollary 2 about the ranges from which we select the parameters Δ_l and Δ_h.

Corollary 2: By jointly considering Theorem 9 and 10, in the incomplete information scenario, the parameters Δ_l and Δ_h should satisfy $F^{-1}\big(1 - \sqrt[S]{1-\theta_{ic}}\big) \leq \Delta_l < \Delta_h \leq \overline{\Delta}_h$, in order to guarantee, with high probability, the existence of at least one participating worker at the corresponding BNE (Theorem 9), and that with high probability Theseus yields a small approximation ratio (Theorem 10).

C. Summary of Parameterization

Thus far, we have finished our discussion of parameterizing Theseus in both the complete (Section VI-A) and incomplete (Section VI-B) information scenario. In summary, in the complete information scenario, if parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ and Δ_t satisfy Condition (16)-(18) and Corollary 1, at the BNE derived in Theorem 1, Theseus satisfies budget feasibility (Theorem 2), individual rationality (Theorem 3), as well as with high probability it has a small approximation ratio (Theorem 5), and with high probability it guarantees that there exist participating workers (Theorem 4). Similarly, in the incomplete information scenario, if we set parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$, Δ_l, and Δ_h according to Condition (22)-(25) and Corollary 2, at the BNE characterized in Theorem 6, Theseus also satisfies budget feasibility (Theorem 7), individual rationality (Theorem 8), as well as with high probability it guarantees that there will be participating workers (Theorem 9), and with high probability it has a small approximation ratio (Theorem 10).

VII. Extensions to MCS Systems With Binary Classification Tasks

In this section, we extend our model, problem formulation, payment mechanism design, and the corresponding analyses to MCS systems with binary classification tasks.

A. Model

Different from Section III-A, the platform in the MCS system considered in this section holds a set of M binary classification tasks (e.g., labeling whether bumps and potholes exist on a piece of road surface [5]), each of which has either 0 or 1 as its true label. We again use $\mathcal{M} = \{1, 2 \cdots, M\}$ to denote the task set. In the rest of Section VII-A, we will only elaborate on where the model is different from its counterpart given in Section III.

We use x_m^{truth} to denote the ground truth label of task m, x_m^s to denote worker s's label on task m, and X_m^{truth} and X_m^s to denote, respectively, the random variables corresponding to x_m^{truth} and x_m^s. Next, we define $e_s \in [\underline{e}_s, \overline{e}_s]$ as the probability that worker s provides a wrong label on task m, i.e., $\Pr(X_m^s \neq X_m^{\text{truth}}) = e_s$. Clearly, the smaller e_s is, the more likely that worker s's sensory data will be close to the ground truths.

In this section, we use e_s as a worker s's strategy in the sensing game induced by the payment mechanism, and use $e_s = \perp$ to denote that worker s chooses to drop out. Thus, a worker s's strategy space is $[\underline{e}_s, \overline{e}_s] \cup \{\perp\}$. Given workers' strategy profile $\mathbf{e} = \{e_1, e_2, \cdots, e_S\}$, a worker s's utility $u_s(\mathbf{e})$ has the same form as Equation (4) except that we substitute δ_s with e_s, and $\boldsymbol{\delta}$ with $\mathbf{e}$.

B. Mathematical Formulation and Proposed Mechanism

In this section, we also take individual rationality, budget feasibility, and low error probability as our design objectives. Next, we introduce the following Lemma 2 that establishes an upper bound for the error probability given workers' strategy profile $\mathbf{e}$.

Lemma 2: Given any $\alpha > 0$, and workers' strategy profile $\mathbf{e}$, we have that

$$\Pr\left(\frac{1}{M}\sum_{m=1}^{M}\left|X_m^* - X_m^{\text{truth}}\right| \geq \alpha\right) \leq \frac{\sum_{s\in\mathcal{S}'} e_s}{\alpha}, \qquad (29)$$

that is, the error probability of a truth discovery algorithm is upper bounded by $\frac{\sum_{s\in\mathcal{S}'} e_s}{\alpha}$.

Proof: Please refer to the supplementary material [43] for the detailed proof. □

Similar to the PMD problem formulated in Section IV, we aim to obtain the payment mechanism $p(\cdot)$ that minimizes $\sum_{s\in\mathcal{S}'} e_s^*$ at the corresponding BNE $\mathbf{e}^*$. The mathematical formulation of designing such payment mechanism, which is referred to as the cPMD problem, is provided in the following optimization program.

cPMD Problem:

$$\min_{p(\cdot)\in\mathcal{P}} \sum_{s\in\mathcal{S}'} e_s^* \qquad (30)$$

$$\text{s.t. } \mathbb{E}_{\mathbf{e}^*_{-s}}\left[u_s(e_s^*, \mathbf{e}^*_{-s})\right] \geq 0, \quad \forall s \in \mathcal{S} \qquad (31)$$

$$\mathbb{E}_{\mathbf{e}^*}\left[\sum_{s\in\mathcal{S}'} p_s\big(\mathbf{X}(\mathbf{e}^*)\big)\right] \leq B \qquad (32)$$

The detailed description of the cPMD problem is omitted, as it could be directly adapted from that of the PMD problem. Next, we propose to use the same payment mechanism as given in Algorithm 2, and provide the corresponding parameterization of the parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ in Section VII-C.

C. Parameterization

To simplify our analysis, we assume that each worker s's cost function is quadratic[5] w.r.t. $e_s \in [\underline{e}_s, \overline{e}_s]$, i.e.,

$$C_s(e_s) = -c_{s,1}e_s^2 + c_{s,2}, \quad \forall e_s \in [\underline{e}_s, \overline{e}_s], \qquad (33)$$

where $c_{s,1}$ and $c_{s,2}$ are positive parameters. Next, we consider the complete information scenario in Section VII-C1, where each worker s's $c_{s,1}$ and $c_{s,2}$ are *a priori* known, and the incomplete information scenario in Section VII-C2, where the platform only knows that $c_{s,1} \in [\underline{c}_1, \overline{c}_1]$ and $c_{s,2} \in [\underline{c}_2, \overline{c}_2]$. In both sections, we assume that the lower bounds of workers' e_s's, i.e., $\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_S$, are i.i.d. random variables within the range $[\underline{e}, \overline{e}]$ with *a priori* PDF $g(\cdot)$.

1) Complete Information Scenario: As each worker s's $c_{s,1}$ and $c_{s,2}$ are known in the complete information scenario, for any $\Lambda_t \in [\underline{e}, \overline{e}]$ and $\overline{e} < 0.5$, we propose to parameterize Theseus with any set of parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ that

[5]Note that the variables of Equation (15) and (33) have rather different meanings. That is, Equation (15) is linear w.r.t. the standard deviation of the difference between worker s's data and the ground truth, whereas Equation (33) is quadratic w.r.t. worker s's error probability for a binary classification task. We leave the investigation of whether our results hold for general forms of cost functions in our future work.

satisfy Condition (34)-(36).

$$\begin{cases} a_s \geq \dfrac{2c_{s,1}\overline{e}}{1-2\overline{e}}, \quad \forall s \in \mathcal{S} & (34) \\ b_s = a_s\Lambda_t - c_{s,1}\Lambda_t^2 + c_{s,2} - a_s(2\Lambda_t - 1)D(\Lambda_t), \\ \qquad \forall s \in \mathcal{S} & (35) \\ \displaystyle\sum_{s=1}^{S} b_s \leq B + 2\big(\underline{e} - \Lambda_t^2\big)\sum_{s=1}^{S} a_s, & (36) \end{cases}$$

where $D(\Lambda_t) = \int_{\underline{e}}^{\Lambda_t} u \frac{g(u)}{\int_{\underline{e}}^{\Lambda_t} g(v)dv} du$. Then, we start our analysis of such parameterization by deriving the BNE of the corresponding sensing game in the following Theorem 11.

Theorem 11: If parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ satisfy Condition (34) and (35) , we have that $\mathbf{e}^* = (e_1^*, e_2^*, \cdots, e_S^*)$, where, for each worker $s \in \mathcal{S}$,

$$e_s^* = \begin{cases} \perp, & \text{if } \underline{e}_s > \Lambda_t \\ \underline{e}_s, & \text{if } \underline{e}_s \leq \Lambda_t, \end{cases} \tag{37}$$

is a BNE of the sensing game in the complete information scenario.

Proof: Please refer to the supplementary material [43] for the detailed proof. □

We next introduce Theorem 12-15 without proofs, as they could be obtained by adapting those of Theorem 2-5. In the following Theorem 12 and 13, we show that Theseus satisfies budget feasibility and individual rationality, if it is parameterized according to Condition (34)-(36).

Theorem 12: Condition (36) ensures that Theseus is budget feasible.

Theorem 13: Theseus is individual rational in MCS systems with binary classification tasks in the complete information scenario.

Next, we discuss our criterion of selecting the parameter Λ_t. We let $\underline{e}_{(1)} = \min\{\underline{e}_1, \underline{e}_2, \cdots, \underline{e}_S\}$, and assume that the CDF $G(\cdot)$ of any $\underline{e}_s$ is invertible with an inverse $G^{-1}(\cdot)$. In the following Theorem 14, we establish a lower bound for Λ_t.

Theorem 14: Given any $\theta_c \in (0,1)$, if $\Lambda_t \geq G^{-1}\big(1 - \sqrt[S]{1-\theta_c}\big)$, then $\Pr\big(\underline{e}_{(1)} \leq \Lambda_t\big) \geq \theta_c$, i.e., the probability that at least one worker chooses to participate at the BNE of the sensing game in MCS systems with binary classification tasks, in the complete information scenario, is no less than the threshold θ_c.

Similar to Theorem 5, we use APP to denote the value of the cPMD problem's objective function guaranteed by Theseus, and we let OPT $= \underline{e}_{(1)}$ which is the objective function's minimum value. In Theorem 15, we derive a probabilistic guarantee on the approximation ratio between APP and OPT.

Theorem 15: In the complete information scenario, given $\alpha_c > 1$ and $\beta_c \in (0,1)$, we have that

$$\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_c\right) \leq \beta_c, \tag{38}$$

if $\Lambda_t \leq \overline{\Lambda}_t$, where $\overline{\Lambda}_t$ is the solution to

$$\overline{\Lambda}_t + \sqrt{-\frac{2}{S\ln\beta_c}}\Big(Q\big(\overline{\Lambda}_t\big)S - \underline{e}\alpha_c\Big) = 0, \tag{39}$$

with $Q\big(\overline{\Lambda}_t\big) = \int_{\underline{e}}^{\overline{\Lambda}_t} u \frac{g(u)}{\int_{\underline{e}}^{\overline{\Lambda}_t} g(v)dv} du$.

Apart from the approximation ratio, Theorem 15 also establishes an upper bound $\overline{\Lambda}_t$ for the parameter Λ_t. Then, we have the following Corollary 3 about the range of Λ_t.

Corollary 3: By jointly considering Theorem 14 and 15, Δ_t should satisfy $G^{-1}\big(1 - \sqrt[S]{1-\theta_c}\big) \leq \Delta_t \leq \overline{\Delta}_t$ in the complete information scenario, so as to guarantee that with high probability there exist participating workers at the corresponding BNE (Theorem 14), and that with high probability Theseus has a small approximation ratio (Theorem 15).

2) Incomplete Information Scenario: In this section, we study the incomplete information scenario, where, instead of the exact values of each worker s's $c_{s,1}$ and $c_{s,2}$, the platform only knows that $c_{s,1} \in [\underline{c}_1, \overline{c}_1]$, and $c_{s,2} \in [\underline{c}_2, \overline{c}_2]$ for each worker s. Then, given any Λ_l and Λ_h with $\underline{e} \leq \Lambda_l < \Lambda_h \leq \overline{e}$, we propose to parameterize Theseus with any set of parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ such that Condition (40)-(43) are satisfied.

$$\begin{cases} a_s \geq \dfrac{2\overline{c}_1\overline{e}}{1-2\overline{e}}, \quad \forall s \in \mathcal{S} & (40) \\ b_s \leq a_s\Lambda_h - \overline{c}_1\Lambda_h^2 + \underline{c}_2 - 2a_s(\Lambda_h - 1)D(\Lambda_h), \\ \qquad \forall s \in \mathcal{S} & (41) \\ b_s \geq a_s\Lambda_l - \underline{c}_1\Lambda_l^2 + \overline{c}_2 - 2a_s(\Lambda_l - 1)D(\Lambda_l), \\ \qquad \forall s \in \mathcal{S} & (42) \\ \displaystyle\sum_{s=1}^{S} b_s \leq B + 2\big(\underline{e} - \Lambda_l^2\big)\sum_{s=1}^{S} a_s, & (43) \end{cases}$$

where $D(\Delta_h) = \int_{\underline{e}}^{\Lambda_h} u \frac{g(u)}{\int_{\underline{e}}^{\Lambda_h} g(v)dv} du$. Next, we introduce the following Theorem 16-20 regarding the analysis of such parameterization, the proofs of which are omitted because they can be adapted from their counterparts in Section VI-B2. Specifically, in Theorem 16, we characterize the BNE guaranteed by such parameterization.

Theorem 16: If parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ satisfy Condition (40)-(42), we have a BNE $\mathbf{e}^* = \big(e_1^*, e_2^*, \cdots, e_S^*\big)$ of the sensing game in the incomplete information scenario, such that, for each worker $s \in \mathcal{S}$,

$$es^* = \begin{cases} \perp, & \text{if } \underline{e}_s > \Lambda_h \\ \underline{e}_s, & \text{if } \underline{e}_s \leq \Lambda_l. \end{cases} \tag{44}$$

Next, Theorem 17 and 18 shows respectively that Theseus is budget feasible and individual rational under the parameterization given by Condition (40)-(43).

Theorem 17: Condition (43) ensures that Theseus is budget feasible in the incomplete information scenario.

Theorem 18: Theseus is individual rational in MCS systems with binary classification sensing tasks in the incomplete information scenario.

Then, in the following Theorem 19, we establish a lower bound for the parameter Λ_l.

Theorem 19: Given any $\theta_{ic} \in (0,1)$, if $\Lambda_l \geq G^{-1}\big(1 - \sqrt[S]{1-\theta_{ic}}\big)$, then $\Pr\big(\underline{e}_{(1)} \leq \Lambda_l\big) \geq \theta_{ic}$, i.e., the probability that at least one worker chooses to participate at the BNE of the sensing game in MCS systems with binary classification tasks, in the incomplete information scenario, is no less than the threshold θ_{ic}.

Similar to Theorem 15, we establish an upper bound for the parameter Λ_h and derive a probabilistic guarantee on the approximation ratio between APP and OPT.

Theorem 20: In the incomplete information scenario, given $\alpha_{ic} > 1$ and $\beta_{ic} \in (0,1)$, we have that

$$\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_{ic}\right) \leq \beta_{ic}, \tag{45}$$

if $\Lambda_h \leq \overline{\Lambda}_h$, where $\overline{\Lambda}_h$ is the solution to

$$\overline{\Lambda}_h + \sqrt{-\frac{2}{S \ln \beta_{ic}}}\Big(Q\big(\overline{\Lambda}_h\big)S - \underline{e}\alpha_{ic}\Big) = 0, \qquad (46)$$

with $Q\big(\overline{\Lambda}_h\big) = \int_{\underline{e}}^{\overline{\Lambda}_h} u \frac{f(u)}{\int_{\underline{e}}^{\overline{\Lambda}_h} g(v)dv} du.$

Combining Theorem 19 and 20, we have the following Corollary 4 on the range from which the parameters Λ_l and Λ_h should be selected.

Corollary 4: By jointly considering Theorem 19 and 20, in the incomplete information scenario, Λ_l and Λ_h should satisfy $G^{-1}\big(1 - \sqrt[S]{1-\theta_{ic}}\big) \leq \Lambda_l < \Lambda_h \leq \overline{\Lambda}_h$, so as to guarantee, with high probability, the existence of at least one participating worker at the corresponding BNE (Theorem 19), and that with high probability Theseus yields a small approximation ratio (Theorem 20).

3) Summary of Parameterization: So far, our parameterization of Theseus in the complete information scenario (Section VII-C1) and incomplete information scenario (Section VII-C2) in MCS systems with binary classification tasks have been finished. In summary, in the complete information scenario, if parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$ and Λ_t satisfy Condition (34)-(36) and Corollary 3, at the BNE derived in Theorem 11, Theseus satisfies budget feasibility (Theorem 12), individual rationality (Theorem 13), as well as with high probability it has a small approximation ratio (Theorem 15), and with high probability it guarantees that there exist participating workers (Theorem 14). Similarly, in the incomplete information scenario, if we set parameters $\{(a_s, b_s)|s \in \mathcal{S}\}$, Λ_l, and Λ_h according to Condition (40)-(43) and Corollary 4, at the BNE characterized in Theorem 16, Theseus also satisfies budget feasibility (Theorem 17), individual rationality (Theorem 18), as well as with high probability it guarantees that there will be participating workers (Theorem 19), and with high probability it has a small approximation ratio (Theorem 20).

VIII. Performance Evaluation

In this section, we introduce the baseline methods, as well as simulation settings and results.

A. Baseline Methods

In our numerical evaluation for MCS systems with continuous tasks, we consider two baseline methods, namely *Max Std* and *Random Std*. In Max Std, each worker s takes strategy $\overline{\delta}_s$, i.e., the maximum standard deviation of the difference between her data and the ground truths. Max Std actually corresponds to the family of payment mechanisms that provide rather insufficient incentives so that workers are only willing to spend little amount of effort. Different from Max Std, in Random Std, each worker s selects her strategy δ_s uniformly at random from the range $[\underline{\delta}_s, \overline{\delta}_s]$. Similarly, in our evaluation for MCS systems with binary classification tasks, we consider the baseline method *Max EP* where each work s takes strategy $\overline{e}_s$, and another baseline method *Random EP* where each worker s's strategy e_s is selected uniformly at random from the range $[\underline{e}_s, \overline{e}_s]$. We compare these baseline methods with the BNEs of the sensing game induced by Theseus, established in Theorem 1, 6, 11, and 16. Note that we do not compare Theseus with existing mechanisms in past literature, because, as indicated in Section II, none of them consider the same scenario as this paper, and thus they are not comparable with Theseus.

TABLE II
Simulation Settings

Setting	$\underline{\delta}_s, \underline{e}_s$	$\overline{\delta}_s, \overline{e}_s$	θ_c, θ_{ic}	α_c, α_{ic}	β_c, β_{ic}	x_m^{truth}	S	M
I, III	$[0.1, 4]$	$[5, 10]$	0.9	5	0.1	$[0, 10]$	$[120, 150]$	30
II, IV	$[0.1, 4]$	$[5, 10]$	0.9	5	0.1	$[0, 10]$	130	$[10, 40]$
V, VI	$[0.05, 0.15]$	$[0.45, 0.5]$	0.9	150	0.2	$\{0, 1\}$	110	$[30, 60]$

B. Simulation Settings

For MCS systems with continuous tasks, we consider setting I-IV in Table II. Among them, setting I and II correspond to the complete information scenario, whereas setting III and IV correspond to the incomplete information scenario. In setting I and II, for each worker s, $\underline{\delta}_s$ is generated uniformly at random from the range $[0.1, 4]$, i.e., $\underline{\delta}_s \sim U[0.1, 4]$. Furthermore, we set $\theta_c = 0.9$, $\alpha_c = 5$, and $\beta_c = 0.1$, and generate $\overline{\delta}_s$ and x_m^{truth} uniformly at random from the range $[5, 10]$ and $[0, 10]$, respectively. In setting I, we fix the number of tasks as $M = 30$ and vary the number of workers S from 120 to 150, whereas in setting II, we fix the number of workers as $S = 130$ and vary the number of tasks M from 10 to 40. Note that the parameter Δ_t is generated uniformly at random from the range $[F^{-1}(1 - \sqrt[S]{1-\theta_c}), \overline{\Delta}_t]$. In setting III and IV for the incomplete information scenario, we generate the parameters $\underline{\delta}_s$, $\overline{\delta}_s$, θ_{ic}, α_{ic}, β_{ic}, x_m^{truth}, S, and M in the same way as in setting I and II, and select Δ_l and Δ_h uniformly at random from the range $[F^{-1}(1 - \sqrt[S]{1-\theta_{ic}}), \overline{\Delta}_h]$ such that $\Delta_l < \Delta_h$.

In all these settings, given δ_s and x_m^{truth}, worker s's data on task m, which is x_m^s, is generated by adding a randomly sampled noise from the distribution $N(0, \delta_s^2)$ to the ground truth x_m^{truth}. Then, workers' data generated by Max Std and Random Std, as well as at the BNEs of the sensing game induced by Theseus, are treated as the inputs to a truth discovery algorithm, respectively, to calculate the estimated ground truths. For MCS systems with binary classification tasks, the parameters $\underline{e}_s$, $\overline{e}_s$, θ_{ic}, α_{ic}, β_{ic}, x_m^{truth}, S, and M are set in the same way as in setting I-IV. Given e_s and x_m^{truth}, we generate x_m^s by setting it to be x_m^{truth} with probability $1-e_s$, and $1 - x_m^{\text{truth}}$ with probability e_s.

In our simulation, the truth discovery algorithm that we implement is the widely adopted CRH [7], which calculates each participating worker s's weight w_s as

$$w_s = \log\left(\frac{\sum_{s' \in \mathcal{S}'}\sum_{m \in \mathcal{M}} |x_m^{s'} - x_m^*|^2}{\sum_{m \in \mathcal{M}} |x_m^s - x_m^*|^2}\right). \qquad (47)$$

C. Simulation Results

In this section, we firstly demonstrate our simulation results regarding the comparison among Max Std, Random Std, and Theseus, in terms of the truth discovery algorithm CRH's MAEs (defined in Definition 6), in Figure 2-5. Note that for each specific worker and task number, we repeatedly generate workers' data, run the truth discovery algorithm CRH, and calculate the corresponding MAE for 10000 times.

In Figure 2 and 3, we plot the means and standard deviations of the MAEs corresponding to Max Std, Random Std, and

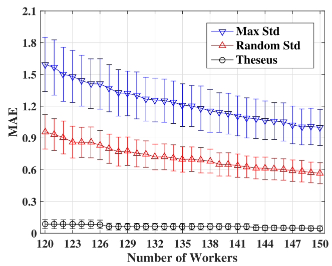

Fig. 2. MAE (setting I).

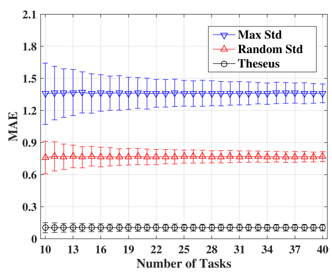

Fig. 3. MAE (setting II).

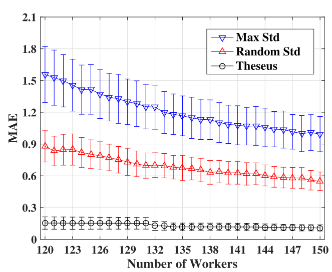

Fig. 4. MAE (setting III).

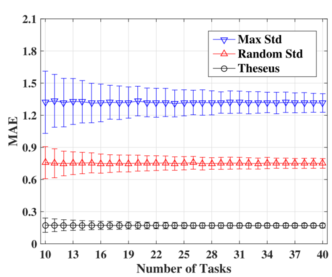

Fig. 5. MAE (setting IV).

Theseus for setting I and II of the complete information scenario. From these two figures, we observe that the means and standard deviations of the MAEs that correspond to Theseus are far less than those that correspond to Max Std and Random Std, which is because Theseus incentivizes workers to exert their maximum effort, so that the standard deviation between each worker's data and the ground truths is minimized. Note that, in Figure 2, the mean of MAE largely decreases as the number of workers increase, because more data that are close to the ground truths will be inputted to CRH with more number of workers. Figure 4 and 5 demonstrate similar trends for the MAEs in setting III and IV of the incomplete information scenario. In Figure 6 and 7,[6] we compare Theseus with Max EP and Random EP in the complete and incomplete information

[6]Note that for ease of presentation, we scale up all standard deviations in Figure 6 and 7 by 2.5 times.

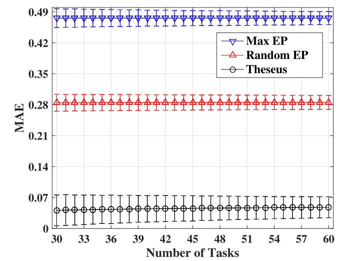

Fig. 6. MAE (setting V).

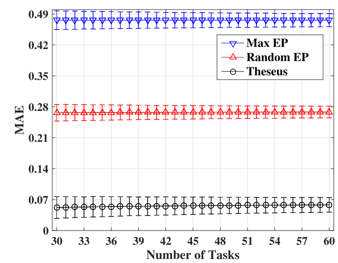

Fig. 7. MAE (setting VI).

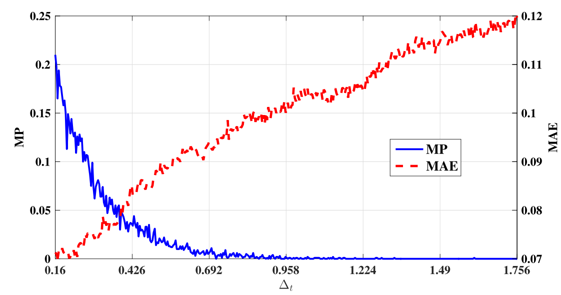

Fig. 8. Trade-off between the MAE and MP.

scenario, respectively, in MCS systems with binary classification tasks. These two figures show that Theseus yields much lower MAEs than MAX EP and Random EP.

Basically, Figure 2-7 indicate collectively that a truth discovery algorithm will return rather inaccurate aggregated results, when a vast majority of the participating workers provide unreliable data. Therefore, our Theseus payment mechanism is highly necessary in order to achieve high aggregation accuracy, even though the platform aggregates workers' data using a state-of-the-art truth discovery algorithm.

In Figure 8, we consider the same parameter settings as in setting I, except that we fix the number of workers as 130, and show the trade-off between the MAE corresponding to Theseus and the *miss probability (MP)*, defined as $\Pr\big(\underline{\delta}_{(1)} > \Delta_t\big)$. Clearly, MP is the probability that no worker will participate at the BNE of the sensing game induced by Theseus. In this figure, the mean of the MAE, as well as the empirical MP obtained by generating $\underline{\delta}_1, \underline{\delta}_2, \cdots, \underline{\delta}_{130}$ according to $U[0.1, 4]$ repeatedly for 10000 times, are plotted. It is clear that although reducing Δ_t causes the decrease of the MAE, it comes at a cost of the increase in the MP. Due to space limit, we omit the figure that shows the trade-off between the MAE and MP in the incomplete information scenario, which has similar trends.

IX. Conclusion

In conclusion, in this paper, we propose a payment mechanism, called Theseus, which is used in pair with a truth discovery algorithm to ensure high aggregation accuracy in

MCS systems where workers may strategically reduce their sensing effort. Theseus tackles workers' strategic behavior, and incentivizes workers to spend their maximum possible effort at the BNE of the induced sensing game among workers. Furthermore, we ensure that Theseus bears other desirable properties, including individual rationality and budget feasibility. The desirable properties of Theseus are validated through theoretical analysis, and extensive simulations.

APPENDIX A
PROOF OF LEMMA 1

Proof: The MAE of a truth discovery algorithm satisfies that

$$\begin{aligned}\frac{1}{M}\sum_{m=1}^{M}\left|X_m^* - X_m^{\text{truth}}\right| &= \frac{1}{M}\sum_{m=1}^{M}\left|\frac{\sum_{s\in\mathcal{S}'} w_s X_m^s}{\sum_{s\in\mathcal{S}'} w_s} - X_m^{\text{truth}}\right| \\ &= \frac{1}{M}\sum_{m=1}^{M}\left|\frac{\sum_{s\in\mathcal{S}'} w_s\big(X_m^s - X_m^{\text{truth}}\big)}{\sum_{s\in\mathcal{S}'} w_s}\right| \\ &\le \frac{1}{M}\frac{\sum_{m=1}^{M}\sum_{s\in\mathcal{S}'} w_s\left|X_m^s - X_m^{\text{truth}}\right|}{\sum_{s\in\mathcal{S}'} w_s} \\ &= \frac{1}{M}\frac{\sum_{s\in\mathcal{S}'} w_s\Big(\sum_{m=1}^{M}\left|X_m^s - X_m^{\text{truth}}\right|\Big)}{\sum_{s\in\mathcal{S}'} w_s} \\ &\le \sum_{s\in\mathcal{S}'}\frac{1}{M}\sum_{m=1}^{M}\left|X_m^s - X_m^{\text{truth}}\right|.\end{aligned}$$

As $X_m^s - X_m^{\text{truth}} \sim N(0,\delta_s^2)$, we have that $\mathbb{E}\big[\big|X_m^s - X_m^{\text{truth}}\big|\big] = \sqrt{\frac{2}{\pi}}\delta_s$. Thus, given any $\alpha > 0$, we have that

$$\begin{aligned}&\Pr\left(\frac{1}{M}\sum_{m=1}^{M}\left|X_m^* - X_m^{\text{truth}}\right| \ge \alpha\right) \\ &\le \Pr\left(\sum_{s\in\mathcal{S}'}\frac{1}{M}\sum_{m=1}^{M}\left|X_m^s - X_m^{\text{truth}}\right| \ge \alpha\right) \\ &\text{(Markov's Inequality)} \\ &\le \frac{\mathbb{E}\big[\sum_{s\in\mathcal{S}'}\frac{1}{M}\sum_{m=1}^{M}\left|X_m^s - X_m^{\text{truth}}\right|\big]}{\alpha} \\ &= \frac{\sum_{s\in\mathcal{S}'}\frac{1}{M}\sum_{m=1}^{M}\mathbb{E}\big[\left|X_m^s - X_m^{\text{truth}}\right|\big]}{\alpha} \\ &= \sqrt{\frac{2}{\pi}}\frac{\sum_{s\in\mathcal{S}'}\delta_s}{\alpha},\end{aligned}$$

which is exactly Inequality (10). □

APPENDIX B
PROOF OF THEOREM 1

Proof: If any worker s chooses to participate, her expected utility, when other workers take strategies $\boldsymbol{\delta}_{-s}^*$, and her reference worker r's strategy δ_r^* is given, can be calculated as

$$\begin{aligned}\mathbb{E}\big[u_s(\delta_s,\boldsymbol{\delta}_{-s}^*)\big|\delta_r^*\big] &= \mathbb{E}\big[p_s\big(\mathbf{X}(\delta_s,\boldsymbol{\delta}_{-s}^*)\big)\big|\delta_r^*\big] - C_s(\delta_s) \\ &= b_s - a_s\mathbb{E}\left[\frac{1}{M}\sum_{m=1}^{M}\big(X_m^s - X_m^r\big)^2\middle|\delta_r^*\right] \\ &\quad + c_{s,1}\delta_s - c_{s,2}.\end{aligned}$$

As $X_m^s - X_m^r = \big(X^{\text{truth}} + N(0,\delta_s^2)\big) - \big(X^{\text{truth}} + N(0,\delta_r^2)\big) = N(0,\delta_s^2) - N(0,\delta_r^2)$, we have $\mathbb{E}\big[(X_m^s - X_m^r)^2\big] = \delta_s^2 + \delta_r^2$. Therefore, we have

$$\mathbb{E}\big[u_s(\delta_s,\boldsymbol{\delta}_{-s}^*)\big|\delta_r^*\big] = b_s - a_s\big(\delta_s^2 + (\delta_r^*)^2\big) + c_{s,1}\delta_s - c_{s,2},$$

and thus,

$$\max\left\{\frac{c_{s,1}}{2a_s},\underline{\delta}_s\right\} = \arg\max_{\delta_s\in[\underline{\delta}_s,\overline{\delta}_s]}\mathbb{E}\big[u_s(\delta_s,\boldsymbol{\delta}_{-s}^*)\big|\delta_r^*\big].$$

That is, regardless of the value of δ_r^*, the strategy $\delta_s \in [\underline{\delta}_s,\overline{\delta}_s]$ that maximizes $\mathbb{E}\big[u_s(\delta_s,\boldsymbol{\delta}_{-s}^*)\big|\delta_r^*\big]$ is the maximum between $\frac{c_{s,1}}{2a_s}$ and $\underline{\delta}_s$. Because of Condition (16), we have that $\underline{\delta}_s \ge \underline{\delta} \ge \frac{c_{s,1}}{2a_s}$. Therefore, if any worker s chooses to participate, her strategy must be $\underline{\delta}_s$, and thus, her expected utility is

$$\begin{aligned}&\mathbb{E}\big[u_s(\underline{\delta}_s,\boldsymbol{\delta}_{-s}^*)\big] \\ &= \mathbb{E}_{\delta_r^*}\Big[\mathbb{E}\big[u_s(\underline{\delta}_s,\boldsymbol{\delta}_{-s}^*)\big|\delta_r^*\big]\Big] \\ &= \mathbb{E}_{\underline{\delta}_r}\Big[\mathbb{E}\big[u_s(\underline{\delta}_s,\boldsymbol{\delta}_{-s}^*)\big|\underline{\delta}_r\big]\Big] \\ &= \mathbb{E}_{\underline{\delta}_r}\big[b_s - a_s\big(\underline{\delta}_s^2 + \underline{\delta}_r^2\big) + c_{s,1}\underline{\delta}_s - c_{s,2}\big] \\ &= b_s - a_s\Big(\underline{\delta}_s^2 + \mathbb{E}_{\underline{\delta}_r}\big[\underline{\delta}_r^2\big|\underline{\delta}_r \le \Delta_t\big]\Big) + c_{s,1}\underline{\delta}_s - c_{s,2} \\ &= \Big(a_s\big(\Delta_t^2 + A(\Delta_t)\big) - c_{s,1}\Delta_t + c_{s,2}\Big) \\ &\quad - \Big(a_s\big(\underline{\delta}_s^2 + A(\Delta_t)\big) - c_{s,1}\underline{\delta}_s + c_{s,2}\Big),\end{aligned}$$

where $A(\Delta_t) = \int_{\underline{\delta}}^{\Delta_t} u^2\frac{f(u)}{\int_{\underline{\delta}}^{\Delta_t} f(v)dv}du$, and the last equality is due to Condition (17). Therefore, for each worker $s \in \mathcal{S}$

$$\begin{cases}\mathbb{E}\big[u_s(\underline{\delta}_s,\boldsymbol{\delta}_{-s}^*)\big] < 0, & \text{if } \underline{\delta}_s > \Delta_t \\ \mathbb{E}\big[u_s(\underline{\delta}_s,\boldsymbol{\delta}_{-s}^*)\big] \ge 0, & \text{if } \underline{\delta}_s \le \Delta_t,\end{cases}$$

and thus, given that other workers take the strategies $\boldsymbol{\delta}_{-s}^*$, worker s will drop out, if $\underline{\delta}_s > \Delta_t$, and will take strategy $\underline{\delta}_s$, if $\underline{\delta}_s \le \Delta_t$. Hence, the strategy profile $\boldsymbol{\delta}^*$ given in Theorem 1 is a BNE of the sensing game. □

APPENDIX C
PROOF OF THEOREM 4

Proof: For any given Δ_t, we have that

$$\begin{aligned}\Pr\big(\underline{\delta}_{(1)} \le \Delta_t\big) &= \Pr\big(\min\{\underline{\delta}_1,\underline{\delta}_2,\cdots,\underline{\delta}_S\} \le \Delta_t\big) \\ &= 1 - \Pr\big(\min\{\underline{\delta}_1,\underline{\delta}_2,\cdots,\underline{\delta}_S\} > \Delta_t\big) \\ &= 1 - \prod_{s=1}^{S}\Pr\big(\underline{\delta}_s > \Delta_t\big) = 1 - \big(1 - F(\Delta_t)\big)^S.\end{aligned}$$

Thus, for any $\theta_c \in (0,1)$, we get $\Delta_t \ge F^{-1}\big(1 - \sqrt[S]{1-\theta_c}\big)$ by setting $1 - \big(1 - F(\Delta_t)\big)^S \ge \theta_c$, which proves Theorem 4. □

APPENDIX D
PROOF OF THEOREM 5

Proof: At the BNE $\boldsymbol{\delta}^*$ given in Theorem 1, we have that

$$\text{APP} = \sum_{s\in\mathcal{S}'}\delta_s^* = \sum_{s\in\mathcal{S}'}\underline{\delta}_s = \sum_{s=1}^{S}\underline{\delta}_s\mathbb{1}_{\{\underline{\delta}_s \le \Delta_t\}},$$

where, for each $s \in \mathcal{S}$, $\mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}}$ is an indicator function with

$$\mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}} = \begin{cases} 0, & \text{if } \underline{\delta}_s > \Delta_t \\ 1, & \text{if } \underline{\delta}_s \leq \Delta_t, \end{cases}$$

and thus, $\underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}} \in [0, \Delta_t]$. Thus, for a fixed $\alpha_c > 1$,

$$\begin{aligned} \Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_c\right) &= \Pr\left(\frac{\sum_{s=1}^{S} \underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}}}{\underline{\delta}_{(1)}} \geq \alpha_c\right) \\ &\leq \Pr\left(\frac{\sum_{s=1}^{S} \underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}}}{\underline{\delta}} \geq \alpha_c\right) \\ &= \Pr\Bigg(\frac{\sum_{s=1}^{S} \left(\underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_t\}} - \mathbb{E}[\underline{\delta}_s | \underline{\delta}_s \leq \Delta_t]\right)}{S} \\ &\geq \frac{\alpha_c \underline{\delta}}{S} - \mathbb{E}\left[\underline{\delta}_1 \middle| \underline{\delta}_1 \leq \Delta_t\right]\Bigg) \\ &\leq \exp\left(-\frac{2\ S^2\left(\frac{\alpha_c \underline{\delta}}{S} - \mathbb{E}[\underline{\delta}_1 | \underline{\delta}_1 \leq \Delta_t]\right)^2}{S \Delta_t^2}\right) \\ &= \exp\left(-\frac{2(\alpha_c \underline{\delta} - R(\Delta_t) S)^2}{S \Delta_t^2}\right), \end{aligned}$$

where the last inequality is because of the Hoeffding's inequality, and $R(\Delta_t) = \int_{\underline{\delta}}^{\Delta_t} u \frac{f(u)}{\int_{\underline{\delta}}^{\Delta_t} f(v) dv} du$. For a fixed $\beta_c \in (0,1)$, by setting $\exp\left(-\frac{2(\alpha_c \underline{\delta} - R(\Delta_t)S)^2}{S \Delta_t^2}\right) \leq \beta_c$, we get that $\Delta_t + \sqrt{-\frac{2}{S \ln \beta_c}}\left(R(\Delta_t) S - \underline{\delta} \alpha_c\right) \leq 0$. Therefore, by setting Δ_t to be no greater than the upper bound $\overline{\Delta}_t$ given in Theorem 5, we have $\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_c\right) \leq \beta_c$. □

APPENDIX E
PROOF OF THEOREM 6

Proof: From Condition (22), we have that $\underline{\delta}_s \geq \underline{\delta} \geq \frac{\overline{c}_1}{2a_s} \geq \frac{c_{s,1}}{2\ a_s}$ for each worker s. By the same reasoning as in the proof of Theorem 1, if a worker s chooses to participate when other workers take strategies $\boldsymbol{\delta}^*_{-s}$, her strategy must be $\underline{\delta}_s$.

Therefore, given that other workers take strategies $\boldsymbol{\delta}^*_{-s}$, worker s's expected utility is

$$\begin{aligned} &\mathbb{E}\left[u_s(\underline{\delta}_s, \boldsymbol{\delta}^*_{-s})\right] \\ &= b_s - a_s\left(\underline{\delta}_s^2 + \mathbb{E}\left[\underline{\delta}_r^2 \middle| \underline{\delta}_r \leq \Delta_h\right]\right) + c_{s,1}\underline{\delta}_s - c_{s,2} \\ &= b_s - a_s\left(\underline{\delta}_s^2 + A(\Delta_h)\right) + c_{s,1}\underline{\delta}_s - c_{s,2}. \end{aligned}$$

Thus, by Condition (23), for any worker s with $\underline{\delta}_s > \Delta_h$,

$$\begin{aligned} \mathbb{E}\left[u_s(\underline{\delta}_s, \boldsymbol{\delta}^*_{-s})\right] \leq\ &\left(a_s\left(\Delta_h^2 + A(\Delta_h)\right) - \overline{c}_1 \Delta_h + \underline{c}_2\right) \\ &- \left(a_s\left(\underline{\delta}_s^2 + A(\Delta_h)\right) - c_{s,1}\underline{\delta}_s + c_{s,2}\right) < 0 \end{aligned}$$

and by Condition (24), for any worker s with $\underline{\delta}_s \leq \Delta_l$,

$$\begin{aligned} \mathbb{E}\left[u_s(\underline{\delta}_s, \boldsymbol{\delta}^*_{-s})\right] \geq\ &\left(a_s\left(\Delta_l^2 + A(\Delta_h)\right) - \underline{c}_1 \Delta_l + \overline{c}_2\right) \\ &- \left(a_s\left(\underline{\delta}_s^2 + A(\Delta_h)\right) - c_{s,1}\underline{\delta}_s + c_{s,2}\right) \geq 0. \end{aligned}$$

Thus, given that other workers take strategies $\boldsymbol{\delta}^*_{-s}$, worker s will drop out, if $\underline{\delta}_s > \Delta_h$, and will take strategy $\underline{\delta}_s$, if $\underline{\delta}_s \leq \Delta_l$. Hence, there exists a BNE $\boldsymbol{\delta}^*$ of the sensing game such that Equation (26) is satisfied for each worker s. □

APPENDIX F
PROOF OF THEOREM 10

Proof: At the BNE of characterized in Theorem 6, we have

$$\text{APP} = \sum_{s \in \mathcal{S}'} \underline{\delta}_s \leq \sum_{s=1}^{S} \underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_h\}},$$

where, for each $s \in \mathcal{S}$, $\mathbb{1}_{\{\underline{\delta}_s \leq \Delta_h\}}$ is an indicator function with

$$\mathbb{1}_{\{\underline{\delta}_s \leq \Delta_h\}} = \begin{cases} 0, & \text{if } \underline{\delta}_s > \Delta_h \\ 1, & \text{if } \underline{\delta}_s \leq \Delta_h. \end{cases}$$

Thus, similar to the proof of Theorem 5, for a fixed $\alpha_{ic} > 1$,

$$\begin{aligned} \Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_{ic}\right) &\leq \Pr\left(\frac{\sum_{s=1}^{S} \underline{\delta}_s \mathbb{1}_{\{\underline{\delta}_s \leq \Delta_h\}}}{\underline{\delta}_{(1)}} \geq \alpha_{ic}\right) \\ &\leq \exp\left(-\frac{2(\alpha_{ic}\underline{\delta} - R(\Delta_h)S)^2}{S \Delta_h^2}\right), \end{aligned}$$

where $R(\Delta_h) = \int_{\underline{\delta}}^{\Delta_h} u \frac{f(u)}{\int_{\underline{\delta}}^{\Delta_h} f(v) dv} du$. Thus, for any fixed $\beta_{ic} \in (0,1)$, by setting $\exp\left(-\frac{2(\alpha_{ic}\underline{\delta} - R(\Delta_h)S)^2}{S \Delta_h^2}\right) \leq \beta_{ic}$, we get $\Delta_h + \sqrt{-\frac{2}{S \ln \beta_{ic}}}\left(R(\Delta_h) S - \underline{\delta}\alpha_{ic}\right) \leq 0$. Therefore, by setting Δ_h to be no greater than the upper bound $\overline{\Delta}_h$ given in Theorem 10, we have that $\Pr\left(\frac{\text{APP}}{\text{OPT}} \geq \alpha_{ic}\right) \leq \beta_{ic}$. □

REFERENCES

[1] H. Jin, L. Su, and K. Nahrstedt, "Theseus: Incentivizing truth discovery in mobile crowd sensing systems," in *Proc. Mobihoc*, 2017, Art. no. 1.
[2] C. Zhou, Y. Gu, S. He, and Z. Shi, "A robust and efficient algorithm for coprime array adaptive beamforming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1099–1112, Feb. 2018.
[3] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "SmartRoad: Smartphone-based crowd sensing for traffic regulator detection and identification," *J. ACM Trans. Sensor Netw.*, vol. 11, no. 4, Dec. 2015, Art. no. 55.
[4] Y. Cheng *et al.*, "AirCloud: A cloud-based air-quality monitoring system for everyone," in *Proc. SenSys*, 2014, pp. 251–265.
[5] J. Eriksson *et al.*, "The pothole patrol: Using a mobile sensor network for road surface monitoring," in *Proc. MobiSys*, 2008, pp. 29–39.
[6] Q. Li *et al.*, "A confidence-aware approach for truth discovery on long-tail data," *J. Proc. VLDB Endowment*, vol. 8, no. 4, pp. 425–436, 2014.
[7] Q. Li *et al.*, "Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation," in *Proc. SIGMOD*, 2014, pp. 1187–1198.
[8] S. Yao *et al.*, "Recursive ground truth estimator for social data streams," in *Proc. IPSN*, Apr. 2016, pp. 1–12.
[9] C. Meng *et al.*, "Truth discovery on crowd sensing of correlated entities," in *Proc. SenSys*, 2015, pp. 169–182.
[10] K. Han *et al.*, "Online pricing for mobile crowdsourcing with multi-minded users," in *Proc. Mobihoc*, 2017, Art. no. 18.
[11] G. Radanovic and B. Faltings, "Incentive schemes for participatory sensing," in *Proc. AAMAS*, 2015, pp. 1081–1089.
[12] G. Radanovic, B. Faltings, and R. Jurca, "Incentives for effort in crowdsourcing using the peer truth serum," *J. ACM Trans. Intell. Syst. Technol.*, vol. 7, no. 4, 2016, Art. no. 48.
[13] M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in *Proc. MobiHoc*, 2015, pp. 157–166.
[14] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "An online pricing mechanism for mobile crowdsensing data markets," in *Proc. Mobihoc*, 2017, Art. no. 26.
[15] X. Gong and N. Shroff, "Incentivizing truthful data quality for quality-aware mobile data crowdsourcing," in *Proc. Mobihoc*, 2018, pp. 161–170.
[16] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Proc. MobiCom*, 2012, pp. 173–184.

[17] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *Proc. ICDCS*, Jun. 2016, pp. 344–353.

[18] H. Jin, L. Su, and K. Nahrstedt, "CENTURION: Incentivizing multi-requester mobile crowd sensing," in *Proc. INFOCOM*, May 2017, pp. 1–9.

[19] Y. Wei, Y. Zhu, H. Zhu, Q. Zhang, and G. Xue, "Truthful online double auctions for dynamic mobile crowdsourcing," in *Proc. INFOCOM*, Apr./May 2015, pp. 2074–2082.

[20] X. Zhang *et al.*, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 12, pp. 3190–3200, Dec. 2014.

[21] Q. Zhang, Y. Wen, X. Tian, X. Gan, and X. Wang, "Incentivize crowd labeling under budget constraint," in *Proc. INFOCOM*, Apr./May 2015, pp. 2812–2820.

[22] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *Proc. MobiHoc*, 2015, pp. 167–176.

[23] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "INCEPTION: Incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *Proc. MobiHoc*, 2016, pp. 341–350.

[24] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems," *IEEE/ACM Trans. Netw.*, vol. 26, no. 5, pp. 2019–2032, Oct. 2018.

[25] H. Jin *et al.*, "Thanos: Incentive mechanism with quality awareness for mobile crowd sensing," *IEEE Trans. Mobile Comput.*, vol. 18, no. 8, pp. 1951–1964, Aug. 2019.

[26] D. Zhao, X.-Y. Li, and H. Ma, "Budget-feasible online incentive mechanisms for crowdsourcing tasks truthfully," *IEEE/ACM Trans. Netw.*, vol. 24, no. 2, pp. 647–661, Apr. 2016.

[27] L. Gao, F. Hou, and J. Huang, "Providing long-term participation incentive in participatory sensing," in *Proc. INFOCOM*, Apr./May 2015, pp. 2803–2811.

[28] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, "Truthful incentive mechanisms for crowdsourcing," in *Proc. INFOCOM*, Apr./May 2015, pp. 2830–2838.

[29] Y. Wen *et al.*, "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Trans. Veh. Technol.*, vol. 64, no. 9, pp. 4203–4214, Sep. 2015.

[30] W. Wang, L. Ying, and J. Zhang, "The value of privacy: Strategic data subjects, incentive mechanisms and fundamental limits," in *Proc. SIGMETRICS*, 2016, pp. 249–260.

[31] S. Ioannidis and P. Loiseau, "Linear regression as a non-cooperative game," in *Proc. WINE*, 2013, pp. 277–290.

[32] R. Cummings, S. Ioannidis, and K. Ligett, "Truthful linear regression," in *Proc. COLT*, 2015, pp. 448–483.

[33] K. Han, H. Huang, and J. Luo, "Posted pricing for robust crowdsensing," in *Proc. MobiHoc*, 2016, pp. 261–270.

[34] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *Proc. INFOCOM*, Apr./May 2014, pp. 745–753.

[35] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in *Proc. MobiHoc*, 2015, pp. 177–186.

[36] L. Jiang, X. Niu, J. Xu, D. Yang, and L. Xu, "Incentivizing the workers for truth discovery in crowdsourcing with copiers," in *Proc. ICDCS*, 2019.

[37] J. Lin, D. Yang, K. Wu, J. Tang, and G. Xue, "A sybil-resistant truth discovery framework for mobile crowdsensing," in *Proc. ICDCS*, 2019.

[38] M. Zhang, L. Yang, X. Gong, S. He, and J. Zhang, "Wireless service pricing competition under network effect, congestion effect, and bounded rationality," *IEEE Trans. Veh. Technol.*, vol. 67, no. 8, pp. 7497–7507, Aug. 2018.

[39] S. Wu *et al.*, "CReam: A smart contract enabled collusion-resistant e-auction," *IEEE Trans. Inf. Forensics Secur.*, vol. 14, no. 7, pp. 1687–1701, Jul. 2019.

[40] Y. Chen *et al.*, "ARMOR: A secure combinatorial auction for heterogeneous spectrum," *IEEE Trans. Mobile Comput.*, vol. 18, no. 10, pp. 2270–2284, Oct. 2019.

[41] B. Zhao and J. Han, "A probabilistic model for estimating real-valued truth from conflicting sources," in *Proc. QDB*, 2012, pp. 1–7.

[42] N. Miller, P. Resnick, and R. Zeckhauser, "Eliciting informative feedback: The peer-prediction method," *Manage. Sci.*, vol. 51, no. 9, pp. 1309–1448, 2005.

[43] *Supplementary Material*. Accessed: Oct. 2019. [Online]. Available: https://shorturl.at/zN135

Haiming Jin (M'19) received the B.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2012, and the Ph.D. degree from the University of Illinois at Urbana-Champaign (UIUC), Urbana, IL, USA, in 2017. Before this, he was a Post-Doctoral Research Associate with the Coordinated Science Laboratory, UIUC. He is currently a tenure-track Assistant Professor with the John Hopcroft Center for Computer Science, Department of Electronic Engineering, Shanghai Jiao Tong University. He is broadly interested in addressing unfolding research challenges in the general areas of urban computing, cyber-physical systems, crowd and social sensing systems, network economics and game theory, reinforcement learning, and mobile and ubiquitous computing.

Baoxiang He is currently pursuing the B.S. degree with the UM-SJTU Joint Institute, Shanghai Jiao Tong University, China. He is also working with Prof. Haiming Jin on mobile crowd sensing and truth discovery.

Lu Su (M'15) received the M.S. degree in statistics and the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign in 2012 and 2013, respectively. He was with the IBM T. J. Watson Research Center and the National Center for Supercomputing Applications. He is currently an Assistant Professor with the Department of Computer Science and Engineering, SUNY Buffalo. His research focuses on the general areas of mobile and crowd sensing systems, Internet of Things, and cyber-physical systems. He is also a member of the ACM. He was a recipient of the NSF CAREER Award, the University at Buffalo Young Investigator Award, the ICCPS17 Best Paper Award, and the ICDCS17 Best Student Paper Award.

Klara Nahrstedt (F'08) received the B.A. degree in mathematics and the M.Sc. degree in numerical analysis from Humboldt University, Berlin, Germany, in 1984 and 1985, respectively, and the Ph.D. degree from the Department of Computer and Information Science, University of Pennsylvania, in 1995.

She was a Research Scientist with the Institute for Informatik, Berlin, in 1990. She is currently a Ralph and Catherine Fisher Full Professor of the Computer Science Department and the Director of the Coordinated Science Laboratory with the University of Illinois at Urbana-Champaign. She is the coauthor of the widely used multimedia books *Multimedia: Computing, Communications, and Applications* (Prentice-Hall, 1995) and *Multimedia Systems* (Springer-Verlag, 2004).

Prof. Nahrstedt is also the member of Leopoldina German National Academy of Sciences. She was a recipient of the IEEE Communication Society Leonard Abraham Award for Research Achievements, University Scholar, the Humboldt Award, the IEEE Computer Society Technical Achievement Award, and the ACM SIG Multimedia Outstanding Technical Achievement Award. She was the General Chair of ACM Multimedia 2006, ACM NOSSDAV 2007, and IEEE Percom 2009. She was elected to serve as the Chair of the ACM SIG Multimedia from 2007 to 2013, and was selected as a member of the Computing Research Association's Computing Community Consortium from 2014 to 2017. She was the Editor-in-Chief of the *Multimedia Systems Journal* from 2000 to 2007, and an Associate Editor of the ACM TOMM from 2005 to 2011. She has been an Associate Editor of the IEEE TRANSACTIONS ON MULTIMEDIA since 2012.

Xinbing Wang (SM'12) received the B.S. degree (Hons.) in automation from Shanghai Jiao Tong University, Shanghai, China, in 1998, the M.S. degree in computer science and technology from Tsinghua University, Beijing, China, in 2001, and the Ph.D. degree with a major in electrical and computer engineering and minor in mathematics from North Carolina State University, Raleigh, in 2006. He is currently a Professor with the Department of Electronic Engineering, Shanghai Jiao Tong University. His research interests include resource allocation and management in mobile and wireless networks, TCP asymptotics analysis, wireless capacity, crosslayer call admission control, asymptotics analysis of hybrid systems, and congestion control over wireless ad hoc and sensor networks. Dr. Wang has been a member of the Technical Program Committees of several conferences, including ACM MobiCom 2012, ACM MobiHoc 2012, and IEEE INFOCOM 2009-2013.