

CENTURION: Incentivizing Multi-Requester Mobile Crowd Sensing

Haiming Jin*, Lu Su[†], Klara Nahrstedt*

*Department of Computer Science, University of Illinois at Urbana-Champaign, IL, USA

[†]Department of Computer Science and Engineering, State University of New York at Buffalo, NY, USA

Email: hjin8@illinois.edu, lusu@buffalo.edu, klara@illinois.edu

Abstract—The recent proliferation of increasingly capable mobile devices has given rise to mobile crowd sensing (MCS) systems that outsource the collection of sensory data to a crowd of participating workers that carry various mobile devices. Aware of the paramount importance of effectively *incentivizing participation* in such systems, the research community has proposed a wide variety of incentive mechanisms. However, different from most of these existing mechanisms which assume the existence of only one data requester, we consider MCS systems with *multiple data requesters*, which are actually more common in practice. Specifically, our incentive mechanism is based on *double auction*, and is able to stimulate the participation of both data requesters and workers. In real practice, the incentive mechanism is typically not an isolated module, but interacts with the *data aggregation mechanism* that aggregates workers' data. For this reason, we propose CENTURION, a novel *integrated framework for multi-requester MCS systems*, consisting of the aforementioned incentive and data aggregation mechanism. CENTURION's incentive mechanism satisfies *truthfulness, individual rationality, computational efficiency*, as well as *guaranteeing non-negative social welfare*, and its data aggregation mechanism generates *highly accurate* aggregated results. The desirable properties of CENTURION are validated through both theoretical analysis and extensive simulations.

I. INTRODUCTION

Recent years have witnessed the rise of mobile crowd sensing (MCS), a newly-emerged sensing paradigm that outsources the collection of sensory data to a crowd of participating users, namely (crowd) workers, who usually carry increasingly capable mobile devices (e.g., smartphones, smartwatches, smart-glasses) with a plethora of on-board sensors (e.g., gyroscope, camera, GPS, compass, accelerometer). Currently, a large variety of MCS systems [1–6] have been deployed that cover almost every aspect of our lives, including healthcare, smart transportation, environmental monitoring, and many others.

To perform the sensing tasks, the participating workers typically consume their own resources such as computing and communicating energy, and expose themselves to potential privacy threats by sharing their personal data. For this reason, a participant would not be interested in participating in the sensing tasks, unless she receives a satisfying reward to compensate her resource consumption and potential privacy breach.

We sincerely thank Professor Julia Chuzhoy for her valuable contribution. We gratefully acknowledge the support of National Science Foundation grants CNS-1330491, and 1566374. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the sponsors.

Therefore, it is necessary to design an effective *incentive mechanism* that can achieve the maximum user participation. Due to the paramount importance of stimulating participation, many incentive mechanisms [7–31] have been proposed by the research community. However, most of these aforementioned past literature assume that there is only one data requester who also serves as the platform in the MCS system. In practice, however, there are usually *multiple data requesters* competing for human resources, who usually outsource worker recruiting to third-party platforms (e.g., Amazon Mechanical Turk [32]) that have already gathered a large number of workers. Therefore, in this paper, we focus on such MCS systems where three parties, including the data requesters, a platform (i.e., a cloud-based central server), as well as a crowd of participating workers co-exist, and aim to develop a *new incentive mechanism that can decide which worker serves which data requester at what price*.

In real practice, the sensory data provided by individual workers are usually quite unreliable due to various factors (e.g., poor sensor quality, lack of sensor calibration, environment noise). Hence, in order to cancel out the possible errors from individual workers, it is highly necessary that the platform utilizes a *data aggregation mechanism* to properly aggregate their noisy and even conflicting data. In an MCS system, the incentive and the data aggregation mechanism are usually not isolated from each other. In fact, the data aggregation mechanism typically interacts with the incentive mechanism, and thus, affects its design and performance. Intuitively, if the platform aggregates workers' data in naive ways (e.g., voting and average) that treat all workers' data equally, the incentive mechanism does not need to distinguish them with respect to their reliability. However, a weighted aggregation method that puts higher weights on more reliable workers is much more desirable, because it shifts the aggregated results towards the data provided by the workers with higher reliability. Accordingly, the incentive mechanism should also incorporate workers' reliability, and selects workers that are more likely to provide reliable data.

Therefore, different from most of the aforementioned existing work [7–31], we propose CENTURION¹, a novel *integrated framework for multi-requester MCS systems*, which

¹The name CENTURION comes from *inCENTivizing mUlti-Requester mobile crOwd seNsing*.

consists of a *weighted data aggregation mechanism* that considers workers' diverse reliability in the calculation of the aggregated results, together with an incentive mechanism that selects workers who potentially will provide more reliable data. Specifically, CENTURION's incentive mechanism is based on *double auction* [33], which involves auctions among not only the workers, but also the data requesters, and is able to incentivize the participation of both data requesters and workers. This paper makes the following contributions.

- Different from existing work, we propose a novel *integrated framework* for *multi-requester* MCS systems, called CENTURION, consisting of a data aggregation and an incentive mechanism. Such an integrated design, which captures the interactive effects between the two mechanisms, is much more complicated and challenging than designing them separately.
- CENTURION's double auction-based incentive mechanism is able to incentivize the participation of both data requesters and workers, and bears many desirable properties, including *truthfulness*, *individual rationality*, *computational efficiency*, as well as *non-negative social welfare*.
- The data aggregation mechanism of CENTURION takes into consideration workers' reliability, and calculates *highly accurate* aggregated results.

In the rest of this paper, we first discuss the past literature that are related to this work in Section II, and introduce the preliminaries in Section III. Then, the design details of CENTURION's data aggregation and incentive mechanism are described in Section IV. In Section V, we conduct extensive simulations to validate the desirable properties of CENTURION. Finally in Section VI, we conclude this paper.

II. RELATED WORK

Aware of the paramount importance of attracting worker participation, the research community has recently developed various incentive mechanisms [7–31] for MCS systems. Among them, game-theoretic incentive mechanisms [7–28], which utilize either auction [10–13, 18–28] or other game-theoretic models [8, 9, 14–17], have gained increasing popularity due to their ability to tackle workers' selfish and strategic behaviors. These mechanisms typically aim to maximize the platform's profit [14–23] or social welfare [9–13], and minimize the platform's payment [7, 8, 24–27] or social cost [28].

Different from most of the aforementioned past literature which assume that there exists only one data requester, we propose a novel incentive mechanism for MCS systems with *multiple data requesters* that compete for human resources. In fact, there do exist several prior work [19, 22, 28] designing incentive mechanisms for the multi-requester scenario. However, they do not provide any *joint design* of the data aggregation and the incentive mechanism as in this paper, which is much more challenging than designing the two mechanisms as isolated modules. Moreover, although similar integrated designs that consider the two mechanisms are proposed in some existing work [25, 26], as previously mentioned, they assume that only one data requester exists in the MCS system.

III. PRELIMINARIES

In this section, we introduce the system overview, reliability level model, auction model, as well as the design objectives.

A. System Overview

CENTURION is an MCS system framework consisting of a cloud-based platform, a set of participating workers, denoted as $\mathcal{W} = \{w_1, \dots, w_N\}$, and a set of requesters, denoted as $\mathcal{R} = \{r_1, \dots, r_M\}$. Each requester $r_j \in \mathcal{R}$ has a sensing task τ_j to be executed by the workers. The set of all requesters' tasks is denoted as $\mathcal{T} = \{\tau_1, \dots, \tau_M\}$. We are specifically interested in the scenario where \mathcal{T} is a set of M different *binary classification tasks* that require workers to locally decide the classes of the events or objects, and report to the platform their local decisions (i.e., the labels of the observed events or objects). Such MCS systems, collecting binary labels from the crowd, constitute a large portion of the currently deployed MCS systems (e.g., congestion detection systems that decide whether or not particular road segments are congested [2], geotagging campaigns that tag whether bumps or potholes exist on specific segments of road surface [1, 3]).

Each task τ_j has a true label $l_j \in \{-1, +1\}$, unknown to the requesters, the platform, and the workers. If a worker w_i is chosen to execute task τ_j , she will provide to the platform a label $l_{i,j}$. We define $\mathbf{I} = [l_{i,j}] \in \{-1, +1, \perp\}^{N \times M}$ as the matrix containing all workers' labels, where $l_{i,j} = \perp$ means that task τ_j is not executed by worker w_i . For every task τ_j , the platform aggregates workers' labels into an aggregated result, denoted as l_j , so as to cancel out the errors from individual workers. The framework of CENTURION is given in Figure 1, and we describe its workflow as follows.

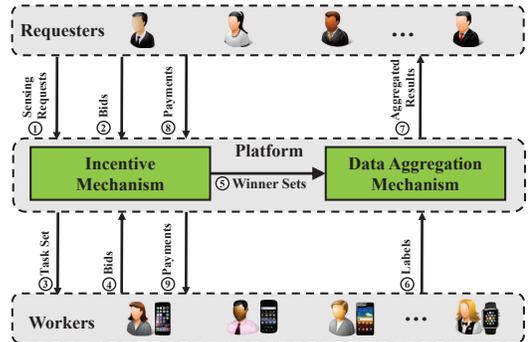


Fig. 1. Framework of CENTURION (where circled numbers represent the order of the events).

- **Incentive Mechanism.** Firstly, in the double auction-based incentive mechanism, each requester r_j submits to the platform a sensing request containing the sensing task τ_j to be executed (step ①), and a bid a_j , the amount she is willing to pay if the task is executed (step ②). Then, the platform announces the set of sensing tasks \mathcal{T} to the workers (step ③). After receiving the task set, every worker w_i sends to the platform the set of tasks she wants to execute, denoted as $\Gamma_i \subseteq \mathcal{T}$, as well as a bid b_i , which is her bidding price for executing them (step ④). Based on received bids, the

platform determines the set of winning requesters $\mathcal{S}_{\mathcal{R}}$, the set of winning workers $\mathcal{S}_{\mathcal{W}}$, as well as the payment p_j^r charged from every winning requester r_j and the payment p_i^w paid to every winning worker w_i (step ⑤). Note that losing requesters' tasks are not executed, and thus, they do not submit any payment. Similarly, losing workers do not receive any payment, as they do not execute any task.

- **Data Aggregation Mechanism.** Next, the platform collects the labels submitted by the winning workers (step ⑥), calculates the aggregated results, and sends them to the winning requesters (step ⑦).

- Finally, the platform charges p_j^r from winning requester r_j (step ⑧), and pays p_i^w to winning worker w_i (step ⑨).

We denote the requesters' and workers' bid profile as $\mathbf{a} = (a_1, \dots, a_M)$ and $\mathbf{b} = (b_1, \dots, b_N)$, respectively. Moreover, the requesters' and workers' payment profile is denoted as $\mathbf{p}^r = (p_1^r, \dots, p_M^r)$ and $\mathbf{p}^w = (p_1^w, \dots, p_N^w)$, respectively.

B. Reliability Level Model

Before worker w_i executes task τ_j , her label about this task can be regarded as a random variable $L_{i,j}$. Then, we define the *reliability level* of a worker in Definition 1.

Definition 1 (Reliability Level). *A worker w_i 's reliability level $\theta_{i,j}$ about task τ_j is defined as the probability that she provides a correct label about this task, i.e.,*

$$\theta_{i,j} = \Pr[L_{i,j} = l_j] \in [0, 1]. \quad (1)$$

Moreover, we denote the workers' reliability level matrix as $\Theta = [\theta_{i,j}] \in [0, 1]^{N \times M}$.

We assume that the platform knows the reliability level matrix Θ *a priori*, and maintains a historical record of it. In practice, the platform could obtain Θ through various approaches. For example, as, in many scenarios, workers tend to have similar reliability levels for similar tasks, the platform could assign to workers some tasks with known labels, and use workers' labels about these tasks to estimate their reliability levels for similar tasks as in [34]. In cases where ground truth labels are not available, Θ can still be effectively inferred from workers' characteristics (e.g., the prices of a worker's sensors, a worker's experience and reputation for similar tasks) using the algorithms proposed in [35], or estimated using the labels previously submitted by workers about similar tasks by the methods in [36, 37].

C. Auction Model

In this paper, we consider the scenario where both requesters and workers are *strategic* and *selfish* that aim to maximize their own utilities. Since CENTURION involves auctions among not only the workers, but also the requesters, we utilize the following *double auction for Multi-rEquester mobiLe crOwd seNsing* (MELON double auction), formally defined in Definition 2, as the incentive mechanism.

Definition 2 (MELON Double Auction). *In a double auction for multi-requester mobile crowd sensing (MELON double auction), each requester r_j obtains a value v_j , if her task τ_j is executed, and bids to the platform a_j , the amount she*

is willing to pay for the execution of her task. Each worker w_i is interested in executing one subset of the tasks, denoted as $\Gamma_i \subseteq \mathcal{T}$, and bids to the platform b_i , her bidding price for executing these tasks. Her actual sensing cost for executing all tasks in Γ_i is denoted as c_i . Both the requesters' values and workers' costs are unknown to the platform.

Then, we define a requester's and worker's utility, as well as the platform's profit in Definition 3, 4, and 5.

Definition 3 (Requester's Utility). *A requester r_j 's utility is defined as*

$$u_j^r = \begin{cases} v_j - p_j^r, & \text{if } r_j \in \mathcal{S}_{\mathcal{R}} \\ 0, & \text{otherwise} \end{cases}. \quad (2)$$

Definition 4 (Worker's Utility). *A worker w_i 's utility is defined as*

$$u_i^w = \begin{cases} p_i^w - c_i, & \text{if } w_i \in \mathcal{S}_{\mathcal{W}} \\ 0, & \text{otherwise} \end{cases}. \quad (3)$$

Definition 5 (Platform's Profit). *The profit of the platform is defined as*

$$u_0 = \sum_{j:r_j \in \mathcal{S}_{\mathcal{R}}} p_j^r - \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}} p_i^w. \quad (4)$$

Based on Definition 3, 4, and 5, we define the social welfare of the MCS system in Definition 6.

Definition 6 (Social Welfare). *The social welfare of the MCS system is defined as*

$$\begin{aligned} u_{\text{social}} &= u_0 + \sum_{i:w_i \in \mathcal{W}} u_i^w + \sum_{j:r_j \in \mathcal{R}} u_j^r \\ &= \sum_{j:r_j \in \mathcal{S}_{\mathcal{R}}} v_j - \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}} c_i. \end{aligned} \quad (5)$$

Clearly, the social welfare is the sum of the platform's profit and all requesters' and workers' utilities.

D. Design Objectives

In this paper, we aim to ensure that CENTURION bears the following advantageous properties.

Since the requesters are strategic and selfish in our model, it is possible that any requester r_j submits a bid a_j that deviates from v_j (i.e., her value for task τ_j). Similarly, any worker w_i might also submit a bid b_i that differs from c_i (i.e., her cost for executing all tasks in Γ_i). Thus, one of our objectives is to design a *truthful* incentive mechanism defined in Definition 7.

Definition 7 (Truthfulness). *A MELON double auction is truthful if and only if bidding v_j and c_i is the dominant strategy for each requester r_j and worker w_i , i.e., bidding v_j and c_i maximizes, respectively, the utility of each requester r_j and worker w_i , regardless of other requesters' and workers' bids.*

By definition 7, we aim to ensure that both requesters and workers bid truthfully to the platform. Apart from truthfulness, another desirable property that we aim to achieve is *individual rationality* defined in Definition 8.

Definition 8 (Individual Rationality). *A MELON double auction is individual rational if and only if no requesters or*

workers receive negative utilities, i.e., we have $u_j^r \geq 0$, and $u_i^w \geq 0$, for every requester r_j and worker w_i , respectively.

Individual rationality is a crucial property to stimulate the participation of both requesters and workers, because it ensures that the charge to a requester is no larger than her value, and a worker's sensing cost is also totally compensated. As mentioned in Section III-A, CENTURION aggregates workers' labels to ensure that the aggregated results have satisfactory accuracy, which is mathematically defined in Definition 9.

Definition 9 (β_j -Accuracy). *A task τ_j is executed with β_j -accuracy if and only if $\Pr[\widehat{L}_j \neq l_j] \leq \beta_j$, where $\beta_j \in (0, 1)$, and \widehat{L}_j denotes the random variable representing the aggregated result for task τ_j .*

By Definition 9, β_j -accuracy ensures that the aggregated result equals to the true label with high probability. Note that, for every task τ_j , β_j is a parameter chosen by the platform, and a smaller β_j implies a stronger requirement for the accuracy.

In short, our objectives are to ensure that the proposed CENTURION framework provides satisfactory *accuracy guarantee* for the aggregated results of all executed tasks, and incentivizes the participation of both requesters and workers in a *truthful and individual rational* manner.

IV. DESIGN DETAILS

In this section, we present the design details of the incentive and data aggregation mechanism of CENTURION.

A. Data Aggregation Mechanism

1) Proposed Mechanism

Although the data aggregation mechanism follows the incentive mechanism in CENTURION's workflow, we introduce it first, as it affects the design of the incentive mechanism.

In order to capture the effect of workers' diverse reliability on the calculation of the aggregated results, CENTURION adopts the following *weighted aggregation method*. That is, the aggregated result \widehat{l}_j for every executed task τ_j is calculated as

$$\widehat{l}_j = \text{sign} \left(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j} l_{i,j} \right), \quad (6)$$

where $\lambda_{i,j} > 0$ is worker w_i 's weight on task τ_j . Furthermore, the function $\text{sign}(x)$ equals to $+1$, if $x \geq 0$, and -1 otherwise.

Intuitively, higher weights should be assigned to workers who are more likely to submit correct labels, which makes the aggregated results closer to the labels provided by more reliable workers. In fact, many state-of-the-art literature [36, 37] utilize such weighted aggregation method to aggregate workers' data. As the weight $\lambda_{i,j}$'s highly affect the accuracy of the aggregated results, we propose, in the following Algorithm 1, the data aggregation mechanism of CENTURION.

Algorithm 1 takes as inputs the reliability level matrix Θ , the workers' label matrix \mathbf{L} , the profile of workers' interested task sets, denoted as $\Gamma = (\Gamma_1, \dots, \Gamma_N)$, the winning requester set $\mathcal{S}_{\mathcal{R}}$, and the winning worker set $\mathcal{S}_{\mathcal{W}}$. Note that a large $\theta_{i,j}$ indicates that a worker w_i has a high reliability level for task τ_j , and any worker w_i with $\theta_{i,j} \leq 0.5$ will not be selected as

Algorithm 1: Data Aggregation Mechanism

Input: $\Theta, \mathbf{L}, \Gamma, \mathcal{S}_{\mathcal{R}}, \mathcal{S}_{\mathcal{W}}$;

Output: $\{\widehat{l}_j | r_j \in \mathcal{S}_{\mathcal{R}}\}$;

1 **foreach** j *s.t.* $r_j \in \mathcal{S}_{\mathcal{R}}$ **do**

2 $\widehat{l}_j \leftarrow \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1) l_{i,j}$;

3 **return** $\{\widehat{l}_j | r_j \in \mathcal{S}_{\mathcal{R}}\}$;

a winner by the incentive mechanism. The aggregated result \widehat{l}_j for each winning requester r_j 's task τ_j is calculated (line 1-2) using Equation (6) with the weight

$$\lambda_{i,j} = 2\theta_{i,j} - 1, \quad \forall r_j \in \mathcal{S}_{\mathcal{R}}, w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i. \quad (7)$$

By Equation (7), we have that $\lambda_{i,j}$, i.e., worker w_i 's weight for task τ_j , increases with $\theta_{i,j}$, which conforms to our intuition that the higher the probability that worker w_i provides a correct label about task τ_j , the more her label $l_{i,j}$ should be counted in the calculation of the aggregated result about this task. We provide the formal analysis about the data aggregation mechanism in Section IV-A2.

2) Analysis

In Theorem 1, we prove that the aggregated results calculated by Algorithm 1 has desirable accuracy guarantee.

Theorem 1. *For each executed task τ_j , the data aggregation mechanism given in Algorithm 1 minimizes the upper bound of the error probability of the aggregated result, i.e., $\Pr[\widehat{L}_j \neq l_j]$ (where \widehat{L}_j is the random variable representing the aggregated result for task τ_j mentioned in Definition 9), and satisfies that*

$$\Pr[\widehat{L}_j \neq l_j] \leq \exp \left(- \frac{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2}{2} \right). \quad (8)$$

Proof. We denote $X_{i,j}$ as the random variable for worker w_i 's weighted label about task τ_j , i.e., $X_{i,j} = \lambda_{i,j} l_j$ with probability $\theta_{i,j}$, and $X_{i,j} = -\lambda_{i,j} l_j$ with probability $1 - \theta_{i,j}$. Then, we define $X_j = \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} X_{i,j}$, and thus, $\mathbb{E}[X_j] = \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \mathbb{E}[X_{i,j}] = \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j} \lambda_{i,j} (2\theta_{i,j} - 1)$.

The error probability of the aggregated result can be calculated as $\Pr[\widehat{L}_j \neq l_j] = \Pr[X_j < 0 | l_j = 1] \Pr[l_j = 1] + \Pr[X_j \geq 0 | l_j = -1] \Pr[l_j = -1]$, and based on the Chernoff-Hoeffding bound, we have

$$\begin{aligned} \Pr[X_j < 0 | l_j = 1] &= \Pr[\mathbb{E}[X_j] - X_j > \mathbb{E}[X_j] | l_j = 1] \\ &\leq \exp \left(- \frac{2(\mathbb{E}[X_j | l_j = 1])^2}{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\lambda_{i,j})^2} \right) \\ &= \exp \left(- \frac{(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j} (2\theta_{i,j} - 1))^2}{2 \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j}^2} \right). \end{aligned}$$

Then, we define the vector $\lambda_j = [\lambda_{i,j}]$ for every executed task τ_j , which contains every $\lambda_{i,j}$ such that $w_i \in \mathcal{S}_{\mathcal{W}}$, and $\tau_j \in \Gamma_i$. Therefore, minimizing the upper bound of $\Pr[X_j < 0 | l_j = 1]$ is equivalent to finding the vector λ_j that maximizes the function $f(\lambda_j)$ defined as

$$f(\lambda_j) = \frac{(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j} (2\theta_{i,j} - 1))^2}{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j}^2}.$$

Based on the Cauchy-Schwarz inequality, we have

$$\begin{aligned} f(\lambda_j) &\leq \frac{(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j}^2)(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2)}{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j}^2} \\ &= \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2, \end{aligned}$$

and equality is achieved if and only if $\lambda_{i,j} \propto 2\theta_{i,j} - 1$. Thus,

$$\Pr[X_j < 0 | l_j = 1] \leq \exp\left(-\frac{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2}{2}\right). \quad (9)$$

Similarly, from the Chernoff-Hoeffding bound, we have

$$\Pr[X_j \geq 0 | l_j = -1] \leq \exp\left(-\frac{(\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j} (2\theta_{i,j} - 1))^2}{2 \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} \lambda_{i,j}^2}\right).$$

The upper bound of $\Pr[X_j > 0 | l_j = -1]$ is also minimized if and only if $\lambda_{i,j} \propto 2\theta_{i,j} - 1$ based on the Cauchy-Schwarz inequality, and we have

$$\Pr[X_j \geq 0 | l_j = -1] \leq \exp\left(-\frac{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2}{2}\right). \quad (10)$$

From Inequality (9) and (10), we have that when $\lambda_{i,j} = 2\theta_{i,j} - 1$, the upper bound of $\Pr[\widehat{L}_j \neq l_j]$ is minimized, and

$$\Pr[\widehat{L}_j \neq l_j] \leq \exp\left(-\frac{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2}{2}\right),$$

which exactly proves Theorem 1. \square

By Theorem 1, we have that the data aggregation mechanism proposed in Algorithm 1 upper bounds the error probability $\Pr[\widehat{L}_j \neq l_j]$ by $\exp\left(-\frac{1}{2} \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2\right)$, which in fact is the minimum upper bound of this probability. Next, we derive Corollary 1, which is directly utilized in our design of the incentive mechanism in Section IV-B.

Corollary 1. *For every executed task τ_j , the data aggregation mechanism proposed in Algorithm 1 satisfies that if*

$$\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2 \geq 2 \ln\left(\frac{1}{\beta_j}\right), \quad (11)$$

then $\Pr[\widehat{L}_j \neq l_j] \leq \beta_j$, i.e., β_j -accuracy is satisfied for this task τ_j , where $\beta_j \in (0, 1)$ is a platform chosen parameter. Moreover, we define $\boldsymbol{\beta}$ as the vector $(\beta_1, \dots, \beta_M)$.

Proof. By setting the upper bound of $\Pr[\widehat{L}_j \neq l_j]$ given in Theorem 1 to be no greater than $\beta_j \in (0, 1)$, we have

$$\exp\left(-\frac{\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2}{2}\right) \leq \beta_j,$$

which is equivalent to

$$\sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2 \geq 2 \ln\left(\frac{1}{\beta_j}\right). \quad (12)$$

Hence, together with Theorem 1, we have that Inequality (12) indicates that $\Pr[\widehat{L}_j \neq l_j] \leq \beta_j$. \square

Corollary 1 gives us a sufficient condition, represented by Inequality (11), that the set of winning workers $\mathcal{S}_{\mathcal{W}}$ selected by the incentive mechanism (proposed in Section IV-B) should satisfy so as to achieve β_j -accuracy for each executed task τ_j .

B. Incentive Mechanism

Now, we introduce the design details of CENTURION's incentive mechanism, including its mathematical formulation, the hardness proof of the formulated integer program, the proposed mechanism, as well as the corresponding analysis.

1) Mathematical Formulation

As mentioned in Section III-C, CENTURION's incentive mechanism is based on the MELON double auction defined in Definition 2. In this paper, we aim to design a MELON double auction that *maximizes the social welfare*, while guaranteeing *satisfactory data aggregation accuracy*. The formal mathematical formulation of its winner selection problem is provided in the following *MELON double auction social welfare maximization (MELON-SWM) problem*.

MELON-SWM Problem:

$$\max \sum_{j:\tau_j \in \mathcal{T}} a_j y_j - \sum_{i:w_i \in \mathcal{W}} b_i x_i \quad (13)$$

$$\text{s.t.} \quad \sum_{i:w_i \in \mathcal{W}, \tau_j \in \Gamma_i} (2\theta_{i,j} - 1)^2 x_i \geq 2 \ln\left(\frac{1}{\beta_j}\right) y_j, \quad \forall \tau_j \in \mathcal{T} \quad (14)$$

$$x_i, y_j \in \{0, 1\}, \quad \forall w_i \in \mathcal{W}, \tau_j \in \mathcal{T} \quad (15)$$

Constants. The MELON-SWM problem takes as inputs the task set \mathcal{T} , the worker set \mathcal{W} , the requesters' and workers' bid profile \mathbf{a} and \mathbf{b} , the profile of workers' interested task sets $\boldsymbol{\Gamma}$, the workers' reliability level matrix $\boldsymbol{\theta}$, and the $\boldsymbol{\beta}$ vector.

Variables. On one hand, the MELON-SWM problem has a vector of M binary variables, denoted as $\mathbf{y} = (y_1, \dots, y_M)$. Any $y_j = 1$ indicates that task τ_j will be executed, and thus, requester r_j is a winning requester (i.e., $r_j \in \mathcal{S}_{\mathcal{R}}$), whereas $y_j = 0$ means $r_j \notin \mathcal{S}_{\mathcal{R}}$. On the other hand, the problem has another vector of N binary variables, denoted as $\mathbf{x} = (x_1, \dots, x_N)$, where $x_i = 1$ indicates that worker w_i is a winning worker (i.e., $w_i \in \mathcal{S}_{\mathcal{W}}$), and $x_i = 0$ means $w_i \notin \mathcal{S}_{\mathcal{W}}$.

Objective function. The objective function satisfies that $\sum_{j:\tau_j \in \mathcal{T}} a_j y_j - \sum_{i:w_i \in \mathcal{W}} b_i x_i = \sum_{j:r_j \in \mathcal{S}_{\mathcal{R}}} a_j - \sum_{i:w_i \in \mathcal{S}_{\mathcal{W}}} b_i$, which is exactly the social welfare defined in Definition 6 based on the requesters' and workers' bids.

Constraints. For each task τ_j , Constraint (14) naturally holds, if $y_j = 0$. When $y_j = 1$, it is equivalent to Inequality (11) given in Corollary 1, which specifies the condition that the set of selected winning workers $\mathcal{S}_{\mathcal{W}}$ should satisfy in order to guarantee β_j -accuracy for task τ_j . To simplify the presentation, we introduce the following notations, namely $q_{i,j} = (2\theta_{i,j} - 1)^2$, $\mathbf{q} = [q_{i,j}] \in [0, 1]^{N \times M}$, $Q_j = 2 \ln\left(\frac{1}{\beta_j}\right)$, and $\mathbf{Q} = [Q_j] \in [0, +\infty)^{M \times 1}$. Thus, Constraint (14) can be simplified as

$$\sum_{i:w_i \in \mathcal{W}, \tau_j \in \Gamma_i} q_{i,j} x_i \geq Q_j y_j, \quad \forall \tau_j \in \mathcal{T}. \quad (16)$$

Besides, we say a task τ_j is *covered* by a solution, if $y_j = 1$.

2) Hardness Proof

We prove the NP-hardness of the MELON-SWM problem by performing a polynomial-time reduction from the 3SAT(5) problem which is formally defined in Definition 10.

Definition 10 (3SAT(5) Problem). *In a 3SAT(5) problem, we are given a set $\mathcal{O} = \{z_1, \dots, z_n\}$ of n Boolean variables, and a collection C_1, \dots, C_m of m clauses. Each clause is an OR of exactly three literals, and every literal is either a variable of \mathcal{O} or its negation. Moreover, every variable participates in exactly 5 clauses. Therefore, $m = \frac{5n}{3}$. Given some constant $0 < \epsilon < 1$, a 3SAT(5) instance φ is a Yes-Instance if there is an assignment to the variables of \mathcal{O} satisfying all clauses, whereas it is a No-Instance (with respect to ϵ), if every assignment to the variables satisfies at most $(1 - \epsilon)m$ clauses. An algorithm A distinguishes between the Yes- and No-instances of the problem, if, given a Yes-Instance, it returns a “YES” answer, and given a No-Instance it returns a “NO” answer.*

Regarding the hardness of the 3SAT(5) problem, we introduce without proof the following well-known Lemma 1, which is a consequence of the PCP theorem [38].

Lemma 1. *There is some constant $0 < \epsilon < 1$, such that distinguishing between the Yes- and No-instances of the 3SAT(5) problem, defined with respect to ϵ , is NP-complete.*

Next, we introduce Theorem 2 and 3 that will be utilized to prove the NP-hardness of the MELON-SWM problem.

Theorem 2. *Any 3SAT(5) instance is polynomial-time reducible to an instance of the MELON-SWM problem.*

Proof. The reduction goes as follows. Assume there is a 3SAT(5) instance φ on n variables and m clauses. We define 3 parameters: $X = \frac{\epsilon m}{100}$ ($0 < \epsilon < 1$), $Y = mnX$, and $Z = mnY$. The exact values of Y and Z are not important. We just need to ensure $Z \gg Y \gg X$. We construct an instance of the MELON-SWM problem corresponding to φ , by defining the task set \mathcal{T} , and the profile of workers’ interested task sets Γ .

Out of the 8 possible assignments to the variables of some clause $C_k \in \varphi$, exactly one does not satisfy C_k . Let A_k be the set of the remaining 7 assignments. We define a set of tasks $\Gamma(C_k, \alpha)$ for each clause C_k and assignment $\alpha \in A_k$, let $\Gamma = [\Gamma(C_k, \alpha)]$ for each clause $C_k \in \varphi$ and assignment $\alpha \in A_k$, set the $q_{i,j}$ value of each worker w_i and task $\tau_j \in \Gamma_i$ as $q_{i,j} = 1$, and set her bid as $b_i = 3 + Y + Z$. We also create a dummy worker w_0 , with $q_0 = 1$, $b_0 = 0$, and Γ_0 being her interested task set. We start with all set $\Gamma(C_k, \alpha)$ ’s being empty, gradually define the tasks, and specify which sets they belong to. The task set \mathcal{T} consists of 4 subsets.

- The 1st subset E_1 contains a task $\tau(z_l, \gamma)$ for each variable $z_l \in \mathcal{O}$ and assignment $\gamma \in \{T, F\}$ to this variable. $\tau(z_l, \gamma)$ belongs to each set $\Gamma(C_k, \alpha)$, such that z_l participates in C_k , and the assignment α to the variables of C_k gives assignment γ to z_l . The Q_j value of the task τ_j corresponding to $\tau(z_l, \gamma)$ is set as $5 -$ the number of the clauses containing z_l , and the value v_j of this task is set as 5.

- The 2nd subset E_2 contains m tasks τ_1, \dots, τ_m . Each $\tau_k \in E_2$ belongs to all sets corresponding to C_k and C_{k+1} , i.e., τ_k belongs to all sets $\{\Gamma(C_k, \alpha) | \alpha \in A_k\} \cup \{\Gamma(C_{k+1}, \alpha') | \alpha' \in A_{k+1}\}$ with the subscripts being modulo m . The Q_k value of each such τ_k is set as 2, and its value v_k is set as Y .
- The 3rd subset E_3 contains a task $\tau(C_k)$ for each clause C_k , and $\tau(C_k)$ belongs to set $\Gamma(C_k, \alpha)$ for each $\alpha \in A_k$. The Q_j value of the task τ_j corresponding to $\tau(C_k)$ is set as 1, and its value v_j is set as Z .
- The 4th subset E_4 contains a single task τ^* , whose Q_j value is set as 1 and value v_j is set as X . The task τ^* only belongs to set Γ_0 .

This finishes the description of the reduction. Clearly, given a 3SAT(5) instance φ , we can construct an instance of the MELON-SWM problem in time polynomial in n . \square

We now analyze the optimal social welfare for an instance of the MELON-SWM problem that corresponds to a 3SAT(5) instance φ , when φ is a Yes- or No-Instance. Note that the following analysis uses the same reduction as in Theorem 2.

Theorem 3. *If the 3SAT(5) instance φ is a Yes-Instance, then there is a solution to the resulting instance of the MELON-SWM problem whose social welfare is X . If φ is a No-Instance, then any solution has social welfare at most 0.*

Proof. Let φ be a Yes-Instance, and A be an assignment to the variables satisfying all clauses. We construct a solution \mathcal{S}' to the MELON-SWM problem. Firstly, we add Γ_0 to \mathcal{S}' . Next, for each clause C_k , we add to \mathcal{S}' the unique set $\Gamma(C_k, \alpha)$, where α is the assignment consistent with A . Then $|\mathcal{S}'| = m$, and the total cost of all sets is $(Y + Z + 3)m$. We now analyze the number of tasks covered by \mathcal{S}' , and their values. Clearly, τ^* is covered by \mathcal{S}' , and it contributes X to the solution value.

- For each clause $C_k \in \varphi$, the unique task $\tau(C_k) \in E_3$ is covered. Thus, all tasks in E_3 are covered, and overall they contribute value mZ to the solution.
- Consider some $\tau_k \in E_2$. \mathcal{S}' contains one set corresponding to C_k and C_{k+1} , respectively. Since τ_k belongs to both these sets, and its Q_k is 2, it is covered. Thus, all tasks in E_2 are covered, and they contribute value mY to the solution.
- Consider some variable $z_k \in \mathcal{O}$, and let $\gamma_k \in \{T, F\}$ be the assignment to z_k under A . If C_k is any clause containing z_k , and $\Gamma(C_k, \alpha)$ is the set that belongs to \mathcal{S}' , then α gives the assignment γ_k to z_k . Thus, for all five clauses containing z_k , the corresponding sets chosen to \mathcal{S}' contain $\tau(z_k, \gamma_k)$, and this task is covered. So the total number of tasks of E_1 covered by \mathcal{S}' is n . Each such task contributes value 5, and the total value contributed by the tasks in E_1 is $5n = 3m$.

Therefore, the overall social welfare of this solution is $X + mZ + mY + 3m - (Z + Y + 3)m = X$.

Assume now that φ is a No-Instance, and let \mathcal{S}' be any solution with positive social welfare. We can assume that $\Gamma_0 \in \mathcal{S}'$, and task τ^* is covered by \mathcal{S}' . We then introduce the following observations, whose proofs are provided in the appendices.

Observation 1. *For every clause C_k of φ , at most one of the sets $\{\Gamma(C_k, \alpha) | \alpha \in A_k\}$ belongs to \mathcal{S}' , and $|\mathcal{S}'| = m$.*

Observation 2. For every variable $z_k \in \mathcal{O}$, at most one of the two tasks $\tau(z_k, T)$ and $\tau(z_k, F)$ is covered by \mathcal{S}' .

We say that a variable $z_k \in \mathcal{O}$ is bad if neither $\tau(z_k, T)$ nor $\tau(z_k, F)$ is covered by \mathcal{S}' ; otherwise it is good. We next show that only a small number of the variables are bad.

Observation 3. There are at most $\frac{\epsilon n}{100}$ bad variables.

Then, we construct the following assignment to the variables of \mathcal{O} . If variable $z_k \in \mathcal{O}$ is good, then there is a unique value $\gamma_k \in \{T, F\}$, such that task $\tau(z_k, \gamma_k)$ is covered by \mathcal{S}' . We then assign z_k the value γ_k . If z_k is bad, we assign it any value arbitrarily. We now claim that the above assignment satisfies more than $(1 - \epsilon)m$ clauses. We say that a clause is bad if it contains a bad variable, and it is good otherwise. Since there are at most $\frac{\epsilon n}{100}$ bad variables, and each variable participates in 5 clauses, the number of bad clauses is at most $\frac{\epsilon n}{20} \leq \frac{3\epsilon m}{100}$. So there are more than $(1 - \epsilon)m$ good clauses. Let C_l be a good clause, and $\Gamma(C_l, \alpha)$ be the set corresponding to C_l that belongs to \mathcal{S}' . Then α is an assignment to the variables of C_l that satisfies C_l , and each variable participating in C_l was assigned a value consistent with α . So clause C_l is satisfied.

To conclude, we have assumed that φ is a No-Instance, and showed that, if the MELON-SWM problem has a solution with non-negative social welfare, there is an assignment to the variables of φ satisfying more than $(1 - \epsilon)m$ of its clauses, which is impossible for a No-Instance. Therefore, if φ is a No-Instance, every solution has social welfare at most 0. \square

Next, we describe Theorem 4 that states the *NP-hardness* and *inapproximability* of the MELON-SWM problem.

Theorem 4. The MELON-SWM problem is NP-hard, and for any factor ϕ , there is no efficient ϕ -approximation algorithm to the MELON-SWM problem.

Proof. Based on Theorem 2, there exists a reduction from any 3SAT(5) problem instance φ to an instance $\mathcal{I}(\varphi)$ of the MELON-SWM problem. From Theorem 3, we have that the optimal solution to $\mathcal{I}(\varphi)$ also gives a solution to φ . That is, if the optimal social welfare of $\mathcal{I}(\varphi)$ is positive, then φ is a Yes-Instance; otherwise, φ is a No-Instance. Together with Lemma 1 stating the NP-completeness of the 3SAT(5) problem, we conclude that the MELON-SWM problem is NP-hard.

In fact, Theorem 2 and 3 give an inapproximability result about the MELON-SWM, as well. Suppose there is an efficient factor- ϕ approximation algorithm \mathcal{A} for the MELON-SWM problem. We can use it to distinguish Yes- and No-instances of the 3SAT(5) problem on $n \gg \phi$ variables. If φ is a Yes-Instance, then the algorithm has to return a solution with positive social welfare for $\mathcal{I}(\varphi)$, and if φ is a No-Instance, then any solution has social welfare at most 0. So algorithm \mathcal{A} distinguishes the Yes- and the No-instances of 3SAT(5), contradicting Lemma 1. \square

3) Proposed Mechanism

Theorem 4 not only shows the NP-hardness of the MELON-SWM problem, but also indicates that there is no efficient algorithm with a guaranteed approximation ratio for it. Therefore, we relax the requirement of provable approximation

ratio, and propose the following MELON double auction that aims to ensure *non-negative social welfare*, instead. Its winner selection algorithm is given in the following Algorithm 2.

Algorithm 2: MELON Double Auction Winner Selection

Input: $\mathcal{T}, \mathcal{R}, \mathcal{W}, \Gamma, \mathbf{a}, \mathbf{b}, \mathbf{q}, \mathbf{Q}$;
Output: $\mathcal{S}_{\mathcal{R}}, \mathcal{S}_{\mathcal{W}}, \mathcal{C}$;
// Initialization
1 $\mathcal{S}_{\mathcal{R}} \leftarrow \emptyset, \mathcal{S}_{\mathcal{W}} \leftarrow \emptyset$;
// Find a feasible cover
2 $\mathcal{C} \leftarrow \text{FC}(\mathcal{T}, \Gamma, \mathbf{q}, \mathbf{Q})$;
3 **foreach** j s.t. $\tau_j \in \mathcal{T}$ **do**
4 $\mathcal{C}_j \leftarrow \{w_i | w_i \in \mathcal{C}, \tau_j \in \Gamma_i\}$;
// Main loop
5 **while** $\max_{j:r_j \in \mathcal{R}} (a_j - \sum_{i:w_i \in \mathcal{C}_j} b_i) \geq 0$ **do**
6 $j^* \leftarrow \arg \max_{j:r_j \in \mathcal{R}} (a_j - \sum_{i:w_i \in \mathcal{C}_j} b_i)$;
7 $\mathcal{S}_{\mathcal{R}} \leftarrow \mathcal{S}_{\mathcal{R}} \cup \{r_{j^*}\}$;
8 $\mathcal{R} \leftarrow \mathcal{R} \setminus \{r_{j^*}\}$;
9 $\mathcal{S}_{\mathcal{W}} \leftarrow \mathcal{S}_{\mathcal{W}} \cup \mathcal{C}_{j^*}$;
10 **foreach** j s.t. $r_j \in \mathcal{R}$ **do**
11 $\mathcal{C}_j \leftarrow \mathcal{C}_j \setminus \mathcal{C}_{j^*}$;
12 **return** $\mathcal{S}_{\mathcal{R}}, \mathcal{S}_{\mathcal{W}}$;

Algorithm 2 takes as inputs the task set \mathcal{T} , the requester set \mathcal{R} , the worker set \mathcal{W} , the profile of workers' interested task sets Γ , the requesters' and workers' bid profile \mathbf{a} and \mathbf{b} , the \mathbf{q} matrix, as well as the \mathbf{Q} vector. Firstly, it initializes the winning requester and worker set as \emptyset (line 1). Then, it calculates a *feasible cover*, denoted by \mathcal{C} , containing the set of workers that make Constraint (16) feasible for each task τ_j given that each $y_j = 1$, by calling another algorithm FC which takes the task set \mathcal{T} , the profile of workers' interested task sets Γ , the \mathbf{q} matrix, and the \mathbf{Q} vector as inputs (line 2). Algorithm FC can be easily implemented in time polynomial in M and N . For example, FC could greedily select each worker w_i into the feasible cover in a decreasing order of the value $\sum_{j:r_j \in \Gamma_i} q_{i,j}$ until all constraints are satisfied. The computational complexity of such FC is $O(N)$. We assume that FC adopts such a greedy approach in the rest of this paper. Note that the specific choice of FC is not important, as long as it returns a feasible cover in polynomial time. Next, for each task τ_j , Algorithm 2 chooses from the feasible cover the set of workers \mathcal{C}_j whose interested task sets contain this task (line 3-4).

Based on \mathcal{C} , the main loop (line 5-11) of the algorithm selects the set of winning requesters and workers that give non-negative social welfare. It executes until $\max_{j:r_j \in \mathcal{R}} (a_j - \sum_{i:w_i \in \mathcal{C}_j} b_i)$, the *maximum marginal social welfare* of including a new requester r_j and the set of workers \mathcal{C}_j into, respectively, the winning requester and worker set, becomes negative (line 5). In each iteration of the main loop, the Algorithm finds first the index j^* of the requester r_{j^*} that provides the *maximum marginal social welfare* (line 6). Next, it includes r_{j^*} into the winning requester set $\mathcal{S}_{\mathcal{R}}$ (line 7), removes r_{j^*} from the requester set \mathcal{R} (line 8), and includes all workers in \mathcal{C}_{j^*} into the winning worker set $\mathcal{S}_{\mathcal{W}}$ (line 9). The last step of the main loop is to remove all workers in \mathcal{C}_{j^*} from \mathcal{C}_j for each task τ_j (line 10). Finally, Algorithm 2 returns

the winning requester and worker set $\mathcal{S}_{\mathcal{R}}$ and $\mathcal{S}_{\mathcal{W}}$ (line 12).

Next, we present the pricing algorithm of the MELON double auction in Algorithm 3.

Algorithm 3: MELON Double Auction Pricing

Input: $\mathcal{T}, \mathcal{R}, \mathcal{W}, \Gamma, \mathbf{a}, \mathbf{b}, \mathbf{q}, \mathbf{Q}, \mathcal{S}_{\mathcal{R}}, \mathcal{S}_{\mathcal{W}}$;
Output: $\mathbf{p}^r, \mathbf{p}^w$;
 // Initialization
1 $\mathbf{p}^r \leftarrow \mathbf{0}, \mathbf{p}^w \leftarrow \mathbf{0}$;
 // Pricing for winning requesters
2 **foreach** j s.t. $r_j \in \mathcal{S}_{\mathcal{R}}$ **do**
3 run Algorithm 2 on $\mathcal{R} \setminus \{r_j\}$ and \mathcal{W} ;
4 $\mathcal{S}'_{\mathcal{R}} \leftarrow$ winning requester set when line 3 stops;
5 **foreach** k s.t. $r_k \in \mathcal{S}'_{\mathcal{R}}$ **do**
6 $p_j^r \leftarrow \min \{p_j^r, \sum_{w_i \in \mathcal{C}'_j} b_i + a_k - \sum_{w_i \in \mathcal{C}'_k} b_i\}$;
7 **if** $\mathcal{C}'_j = \emptyset$ **then**
8 $p_j^r \leftarrow \min\{p_j^r, 0\}$;
 // Pricing for winning workers
9 **foreach** i s.t. $w_i \in \mathcal{S}_{\mathcal{W}}$ **do**
10 run Algorithm 2 on \mathcal{R} and $\mathcal{W} \setminus \{w_i\}$;
11 $\mathcal{S}'_{\mathcal{R}} \leftarrow$ winning requester set when line 10 stops;
12 **foreach** k s.t. $w_i \in \mathcal{C}'_k$ and $r_k \in \mathcal{S}'_{\mathcal{R}}$ **do**
13 sort requesters according to the decreasing order of
 $a_j - \sum_{i: w_i \in \mathcal{C}'_j} b_i$;
14 $f \leftarrow$ index of the first requester with $w_i \notin \mathcal{C}'_j$;
15 **if** $r_f \in \mathcal{S}'_{\mathcal{R}}$ **then**
16 $p_i^w \leftarrow \max \{p_i^w, a_k - \sum_{w_h \in \mathcal{C}'_k} b_h - (a_f - \sum_{w_h \in \mathcal{C}'_f} b_h)\}$;
17 **else**
18 $p_i^w \leftarrow \max \{p_i^w, a_k - \sum_{w_h \in \mathcal{C}'_k} b_h\}$;
19 **return** $\mathbf{p}^r, \mathbf{p}^w$;

Apart from the same inputs to Algorithm 2, Algorithm 3 also takes as inputs the winning requester and worker set $\mathcal{S}_{\mathcal{R}}$ and $\mathcal{S}_{\mathcal{W}}$, outputted by Algorithm 2. Firstly, Algorithm 3 initializes the requesters' and workers' payment profile as zero vectors (line 1). Then, it calculates the payment p_j^r charged from each winning requester (line 2-8). For each $r_j \in \mathcal{S}_{\mathcal{R}}$, Algorithm 2 is executed on the worker set \mathcal{W} and requester set \mathcal{R} except requester r_j (line 3). Next, it sets $\mathcal{S}'_{\mathcal{R}}$ as the winning requester set when line 3 stops (line 4). For each $r_k \in \mathcal{S}'_{\mathcal{R}}$, Algorithm 3 finds the minimum bid $a_{j,k}$ for requester r_j to replace r_k as the winner. To achieve this, $a_{j,k}$ should satisfy $a_{j,k} - \sum_{w_i \in \mathcal{C}'_j} b_i = a_k - \sum_{w_i \in \mathcal{C}'_k} b_i$, which is equivalent to $a_{j,k} = \sum_{w_i \in \mathcal{C}'_j} b_i + a_k - \sum_{w_i \in \mathcal{C}'_k} b_i$. Note that $\mathcal{C}'_1, \dots, \mathcal{C}'_M$ denote the sets $\mathcal{C}_1, \dots, \mathcal{C}_M$ when the specific requester r_k is selected into $\mathcal{S}'_{\mathcal{R}}$. If \mathcal{C}'_j is not empty, the minimum value among these $a_{j,k}$'s is chosen as the payment p_j^r (line 5-6); otherwise, it is further compared with 0 (line 7-8), since requester r_j could win, in this case, as long as her bid is non-negative.

Next, Algorithm 3 derives the payment p_i^w to each winning worker w_i (line 9-18). Similar to line 3, Algorithm 2 is executed on the requester set \mathcal{R} and worker set \mathcal{W} except worker w_i (line 10), and $\mathcal{S}'_{\mathcal{R}}$ is set as the winning requester set when line 10 stops (line 11). In the rest of the algorithm, we also use $\mathcal{C}'_1, \dots, \mathcal{C}'_M$ to denote the sets $\mathcal{C}_1, \dots, \mathcal{C}_M$ when the specific requester r_k is selected into $\mathcal{S}'_{\mathcal{R}}$. For each set

\mathcal{C}'_k such that w_i belongs to \mathcal{C}'_k and r_k belongs to $\mathcal{S}'_{\mathcal{R}}$, the algorithm calculates the maximum bid $b_{i,k}$ for worker w_i to be selected as a winner at this point (line 12-18). The calculation firstly sorts requesters in the decreasing order of their marginal social welfare, i.e., $a_j - \sum_{i: w_i \in \mathcal{C}'_j} b_i$ (line 13), and finds the index f of the first the requester in this order such that w_i does not belong to \mathcal{C}'_f (line 14). If r_f is a winning requester in $\mathcal{S}'_{\mathcal{R}}$, then $b_{i,k}$ should satisfy $a_k - (\sum_{w_h \in \mathcal{C}'_k} b_h + b_{i,k}) = a_f - \sum_{w_h \in \mathcal{C}'_f} b_h$, which is equivalent to $b_{i,k} = a_k - \sum_{w_h \in \mathcal{C}'_k} b_h - (a_f - \sum_{w_h \in \mathcal{C}'_f} b_h)$; otherwise, $b_{i,k}$ should satisfy $a_k - (\sum_{w_h \in \mathcal{C}'_k} b_h + b_{i,k}) = 0$, which is equivalent to $b_{i,k} = a_k - \sum_{w_h \in \mathcal{C}'_k} b_h$. Then, the maximum value among these $b_{i,k}$'s are chosen as the payment p_i^w (line 15-18). Finally, Algorithm 3 returns the requesters' and workers' payment profile \mathbf{p}^r and \mathbf{p}^w (line 19).

4) Analysis of the Proposed Mechanism

In this section, we prove several desirable properties of our MELON double auction, described in Algorithm 2 and 3. Firstly, we show its truthfulness in Theorem 5.

Theorem 5. *The proposed MELON double auction is truthful.*

Proof. We prove the truthfulness of the MELON double auction by showing that it satisfies the properties of *monotonicity* and *critical payment*.

- **Monotonicity.** The algorithm FC called by Algorithm 2 is independent of the requesters' and workers' bids, and winners are selected based on a decreasing order of the value $a_j - \sum_{i: w_i \in \mathcal{C}'_j} b_i$. Thus, if a requester r_j wins by bidding a_j , she will also win the auction by bidding any $a'_j > a_j$. Similarly, if a worker w_i wins by bidding b_i , she will win the auction, as well, if her bid takes any value $b'_i < b_i$.

- **Critical payment.** Algorithm 3 in fact pays every winning requester and worker the infimum and supremum of her bid, respectively, that can make her a winner.

As proved in [39], these two properties make an auction truthful, i.e., each requester r_j maximizes her utility by bidding v_j , and each worker w_i maximizes her utility by bidding c_i . Therefore, the MELON double auction is truthful. \square

Next, we show that the proposed MELON double auction satisfies individual rationality in Theorem 6.

Theorem 6. *The proposed MELON double auction is individual rational.*

Proof. By Definition 3 and 4, losers of the MELON double auction receive zero utilities. From Theorem 5, every winning requester r_j bids v_j , and every winning worker w_i bids c_i to the platform. Moreover, they are paid, respectively, the infimum and supremum of the bid for them to win the auction. Therefore, it is guaranteed that all requesters and workers receive non-negative utilities, and thus the proposed MELON double auction is individual rational. \square

In Theorem 7, we prove that the proposed MELON double auction has a polynomial-time computational complexity.

Theorem 7. *The computational complexity of the proposed MELON double auction is $O(M^3N + M^2N^2)$.*

Proof. As mentioned in Section IV-A1, the algorithm FC (line 2) in Algorithm 2 takes a greedy approach, and has a computational complexity of $O(N)$. Line 3-4 of Algorithm 2 that find the sets $\mathcal{C}_1, \dots, \mathcal{C}_M$ terminate at most after MN steps. Next, the main loop (line 5-11) terminates after M iterations in worst case. Within each iteration, finding the index of the requester that provides the maximum marginal social welfare (line 6) takes $O(M)$ time, and updating the sets $\mathcal{C}_1, \dots, \mathcal{C}_M$ takes $O(MN)$ time. Therefore, the computational complexity of the main loop is $O(MN)$, and thus, that of Algorithm 2 is $O(M^2N)$ overall. After Algorithm 2, our MELON double auction executes its pricing algorithm described by Algorithm 3, where the loop for requester pricing (line 1-8) terminates in worst case after M iterations. Clearly, the computational complexity of each iteration of the loop is dominated by the execution of Algorithm 2 in line 3. Therefore, the requester pricing (line 1-8) in Algorithm 3 takes $O(M^3N)$ time. Following a similar method of analysis, we can conclude that the worker pricing in Algorithm 3 takes $O(M^2N^2)$ time. Hence, the computation complexity of Algorithm 3, as well as that of the overall MELON double auction is $O(M^3N + M^2N^2)$. \square

Finally, we show in Theorem 8 that our MELON double auction guarantees non-negative social welfare, as required.

Theorem 8. *The MELON double auction guarantees non-negative social welfare.*

Proof. Clearly, in the winner selection algorithm described by Algorithm 2, a requester r_j and the workers in \mathcal{C}_j could be selected as winners, only if the corresponding marginal social welfare $a_j - \sum_{i:w_i \in \mathcal{C}_j} b_i$ is non-negative (line 5). Thus, as the overall social welfare given by Algorithm 2 is the sum of the aforementioned marginal social welfare of every iteration where new winners are selected, the MELON double auction guarantees non-negative social welfare. \square

V. PERFORMANCE EVALUATION

In this section, we introduce the baseline methods, simulation settings, as well as simulation results of the performance evaluation about our proposed CENTURION framework.

A. Baseline Methods

In our evaluation of the incentive mechanism, the first baseline auction is the Marginal Social Welfare greedy (MSW-Greedy) double auction. As in Algorithm 2, it also initializes the winner sets as \emptyset , executes the algorithm FC to obtain a feasible cover \mathcal{C} , and chooses from \mathcal{C} the set \mathcal{C}_j containing each worker w_i such that $\tau_j \in \Gamma_i$ for each task τ_j . Different from the MELON double auction, it sorts requesters in a decreasing order of their marginal social welfare, i.e., the value $a_j - \sum_{i:w_i \in \mathcal{C}_j} b_i$ for each requester r_j . Then, it selects the requester r_j and the set of workers in \mathcal{C}_j as winners until the marginal social welfare becomes negative. Its pricing algorithm is the same as that of the MELON double auction. Clearly, the MSW-Greedy double auction is truthful and individual rational. Another baseline auction is the one that initializes the feasible cover \mathcal{C} as the entire worker set \mathcal{W} ,

which we call AIR double auction. The rest of its winner selection, as well as the entire pricing algorithm is the same as those of our MELON double auction. It is easily provable that the AIR double auction is also truthful and individual rational.

Furthermore, we compare our weighted data aggregation mechanism with a mean aggregation mechanism, which outputs +1 as the aggregated result for a task if the mean of workers' labels about this task is non-negative, and outputs -1, otherwise. Another baseline aggregation mechanism that we consider is the median aggregation that takes the median of workers' labels about a task as its aggregated result.

B. Simulation Settings

Setting	v_j	c_i	$\theta_{i,j}$	β_j	$ \Gamma_i^* $	N	M
I	[10, 20]	[5, 15]	[0, 1]	[0.05, 0.1]	[15, 20]	[90, 150]	60
II	[10, 20]	[5, 15]	[0, 1]	[0.05, 0.1]	[15, 20]	60	[20, 80]

TABLE I
SIMULATION SETTINGS

The parameter settings in our simulation are given in Table I. Specifically, parameters v_j , c_i , $\theta_{i,j}$, β_j , and $|\Gamma_i^*|$ are sampled uniformly at random from the intervals given in Table I. The worker w_i 's true interested task set Γ_i^* contains $|\Gamma_i^*|$ tasks that are randomly selected from the task set \mathcal{T} . In setting I, we fix the number of requesters as 60 and vary the number of workers from 90 to 150, whereas we fix the number of workers as 60 and vary the number of requesters from 20 to 80 in setting II.

C. Simulation Results

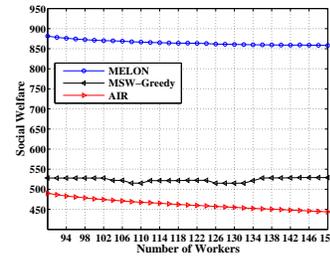


Fig. 2. Social welfare (setting I)

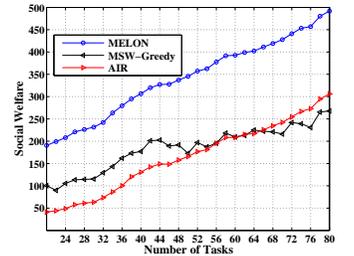


Fig. 3. Social welfare (setting II)

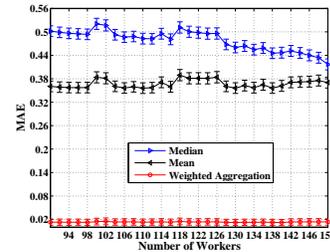


Fig. 4. MAE (setting I)

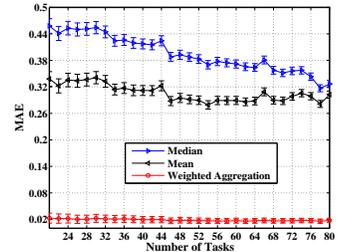


Fig. 5. MAE (setting II)

In Figure 2 and 3, we compare the social welfare generated by our MELON double auction with those of the two baseline auctions. These two figures show that our MELON double auction generates social welfare far more than the MSW-Greedy and AIR double auction under both setting I and II.

We evaluate CENTURION's accuracy guarantee in setting I and II with a minor change of the parameter β_j , i.e., β_j for each task τ_j is fixed as 0.05 to simplify presentation. We

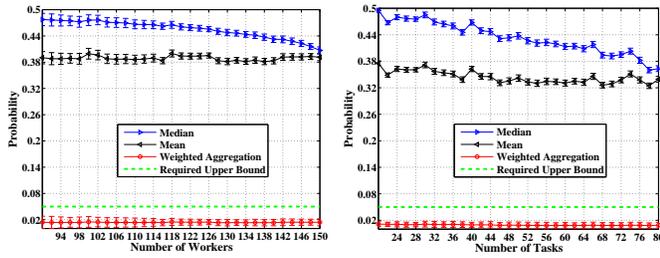


Fig. 6. Error probability (setting I) Fig. 7. Error probability (setting II)

compare the mean absolute error (MAE) for all tasks, which is defined as $MAE = \frac{1}{M} \sum_{j: \tau_j \in \mathcal{T}} |\hat{l}_j - l_j|$, of our weighted aggregation mechanism proposed in Algorithm 1 with those of the mean and median aggregation. The simulation for each combination of worker and requester number is repeated for 50000 times, and we plot the means and standard deviations of the MAEs in Figure 4 and 5. From these two figures, we observe that the MAE of our weighted aggregation mechanism is far less than those of the mean and median aggregation. Then, we show our simulation results about $\Pr[|\hat{l}_j - l_j|]$, referred to as task τ_j 's error probability (EP). After 50000 repetitions of the simulation for any given combination of worker and requester number, empirical values of the EPs are calculated, and the means and standard deviations of the empirical EPs are plotted in Figure 6 and 7. These two figures show that the empirical EPs are less than the required upper bound β_j and far less than those of the mean and median aggregation.

VI. CONCLUSION

In this paper, we propose CENTURION, a novel integrated framework for multi-requester MCS systems, consisting of a double auction-based incentive mechanism that stimulates the participation of both requesters and workers, and a data aggregation mechanism that aggregates workers data. Its incentive mechanism bears many desirable properties including truthfulness, individual rationality, computational efficiency, as well as non-negative social welfare, and its data aggregation mechanism generates highly accurate aggregated results.

REFERENCES

- [1] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: Rich monitoring of road and traffic conditions using mobile smartphones," in *SenSys*, 2008.
- [2] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: Accurate, energy-aware road traffic delay estimation using mobile phones," in *SenSys*, 2009.
- [3] J. Eriksson, L. Girod, B. Hull, R. Newton, S. Madden, and B. Hari, "The pothole patrol: using a mobile sensor network for road surface monitoring," in *MobiSys*, 2008.
- [4] "Myheartmap," <http://www.med.upenn.edu/myheartmap/>.
- [5] S. Hu, L. Su, H. Liu, H. Wang, and T. F. Abdelzaher, "Smartroad: Smartphone-based crowd sensing for traffic regulator detection and identification," in *TOSN*, 2015.
- [6] Y. Cheng, X. Li, Z. Li, S. Jiang, Y. Li, J. Jia, and X. Jiang, "Aircloud: A cloud-based air-quality monitoring system for everyone," in *SenSys*, 2014.
- [7] K. Han, H. Huang, and J. Luo, "Posted pricing for robust crowdsensing," in *MobiHoc*, 2016.

- [8] H. Xie, J. Lui, J. W. Jiang, and W. Chen, "Incentive mechanism and protocol design for crowdsourcing systems," in *Allerton*, 2014.
- [9] M. H. Cheung, R. Southwell, F. Hou, and J. Huang, "Distributed time-sensitive task selection in mobile crowdsensing," in *MobiHoc*, 2015.
- [10] L. Gao, F. Hou, and J. Huang, "Providing long-term participation incentive in participatory sensing," in *INFOCOM*, 2015.
- [11] H. Jin, L. Su, D. Chen, K. Nahrstedt, and J. Xu, "Quality of information aware incentive mechanisms for mobile crowd sensing systems," in *MobiHoc*, 2015.
- [12] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen, "Quality-driven auction based incentive mechanism for mobile crowd sensing," in *TVT*, 2014.
- [13] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: Online incentive mechanisms with budget constraint," in *INFOCOM*, 2014.
- [14] Y. Chen, B. Li, and Q. Zhang, "Incentivizing crowdsourcing systems with network effects," in *INFOCOM*, 2016.
- [15] S. He, D.-H. Shin, J. Zhang, and J. Chen, "Toward optimal allocation of location dependent tasks in crowdsensing," in *INFOCOM*, 2014.
- [16] T. Luo, S. S. Kanhere, H.-P. Tan, F. Wu, and H. Wu, "Crowdsourcing with tulloct contests: A new perspective," in *INFOCOM*, 2015.
- [17] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *INFOCOM*, 2012.
- [18] Q. Zhang, Y. Wen, X. Tian, X. Gan, and X. Wang, "Incentivize crowd labeling under budget constraint," in *INFOCOM*, 2015.
- [19] X. Zhang, G. Xue, R. Yu, D. Yang, and J. Tang, "Truthful incentive mechanisms for crowdsourcing," in *INFOCOM*, 2015.
- [20] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing," in *Mobicom*, 2012.
- [21] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," in *TPDS*, 2014.
- [22] H. Zhang, B. Liu, H. Susanto, G. Xue, and T. Sun, "Incentive mechanism for proximity-based mobile crowd service systems," in *INFOCOM*, 2016.
- [23] Y. Wei, Y. Zhu, H. Zhu, Q. Zhang, and G. Xue, "Truthful online double auctions for dynamic mobile crowdsourcing," in *INFOCOM*, 2015.
- [24] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *INFOCOM*, 2013.
- [25] H. Jin, L. Su, H. Xiao, and K. Nahrstedt, "Inception: incentivizing privacy-preserving data aggregation for mobile crowd sensing systems," in *MobiHoc*, 2016.
- [26] H. Jin, L. Su, B. Ding, K. Nahrstedt, and N. Borisov, "Enabling privacy-preserving incentives for mobile crowd sensing systems," in *ICDCS*, 2016.
- [27] J. Wang, J. Tang, D. Yang, E. Wang, and G. Xue, "Quality-aware and fine-grained incentive mechanisms for mobile crowdsensing," in *ICDCS*, 2016.
- [28] Z. Feng, Y. Zhu, Q. Zhang, L. Ni, and A. Vasilakos, "Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *INFOCOM*, 2014.
- [29] M. Karaliopoulos, I. Koutsopoulos, and M. Titsias, "First learn then earn: Optimizing mobile crowdsensing campaigns through data-driven user profiling," in *MobiHoc*, 2016.
- [30] L. Pu, X. Chen, J. Xu, and X. Fu, "Crowdlet: Optimal worker recruitment for self-organized mobile crowdsourcing," in *INFOCOM*, 2016.
- [31] D. Peng, F. Wu, and G. Chen, "Pay as how well you do: A quality based incentive mechanism for crowdsensing," in

MobiHoc, 2015.

- [32] “Amazon mechanical turk,” <https://www.mturk.com//>.
- [33] R. P. McAfee, “A dominant strategy double auction,” *Journal of economic Theory*, vol. 56, no. 2, pp. 434–450, 1992.
- [34] D. Oleson, A. Sorokin, G. P. Laughlin, V. Hester, J. Le, and L. Biewald, “Programmatic gold: Targeted and scalable quality assurance in crowdsourcing,” in *HCOMP*, 2011.
- [35] H. Li, B. Zhao, and A. Fuxman, “The wisdom of minority: Discovering and targeting the right group of workers for crowdsourcing,” in *WWW*, 2014.
- [36] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, and J. Han, “Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation,” in *SIGMOD*, 2014.
- [37] C. Meng, W. Jiang, Y. Li, J. Gao, L. Su, H. Ding, and Y. Cheng, “Truth discovery on crowd sensing of correlated entities,” in *SenSys*, 2015.
- [38] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy, “Proof verification and the hardness of approximation problems,” in *Journal of the ACM*, 1998.
- [39] L. Blumrosen and N. Nisan, “Combinatorial auctions,” *Algorithmic Game Theory*, 2007.

task of E_4 contributes X . The total value of all tasks covered is then at most $X + 3m - \frac{\epsilon n}{20} + Zm + Ym$, while the total cost of all sets is $m(Z + Y + 3)$. Therefore, the social welfare is $X - \frac{\epsilon n}{20} < 0$, a contradiction. \square

APPENDIX A

PROOF OF OBSERVATION 1

Proof. Let \mathcal{R} be the set of all clauses C_k , such that \mathcal{S}' contains any set of the form $\Gamma(C_k, \alpha)$, and assume that $|\mathcal{R}| = t$. Then the number of tasks of E_3 covered by \mathcal{S}' is exactly t , and they contribute value tZ to the social welfare. All remaining tasks may contribute at most $\frac{Z}{2}$ value to the social welfare. Since the cost of every set in \mathcal{S}' is at least Z , in order for the final social welfare to be non-negative, $|\mathcal{S}'| = t$ must hold, and so \mathcal{S}' contains at most one set corresponding to every clause.

Next, we prove that $|\mathcal{S}'| = m$. Let t be the number of tasks of E_1 covered by the solution. From previous discussions, $|\mathcal{S}'| = t$. Assume for contradiction that $|\mathcal{S}'| < m$. Then the number of task of E_2 covered by \mathcal{S}' is at most $t - 1$. Therefore, the total value of all tasks covered by the solution is upper bounded by $X + tZ + (t - 1)Y + 10n = X + tZ + tY + 6m - Y$, while the total cost of all sets in \mathcal{S}' is $(Z + Y + 3)t$. Therefore, the total profit is at most $X + 6m - Y - 3t < 0$. \square

APPENDIX B

PROOF OF OBSERVATION 2

Proof. Assume for contradiction that for some $z_k \in \mathcal{O}$, both $\tau(z_k, T)$ and $\tau(z_k, F)$ are covered by \mathcal{S}' . Recall that there are exactly five clauses containing the variable z_k . Both tasks only belong to sets corresponding to these five clauses, and each such set contains exactly one of the two tasks. Since the Q_j value of each such task is 5, and for each clause C_k exactly one of its corresponding set belongs to \mathcal{S}' , it is impossible that both $\tau(z_k, T)$ and $\tau(z_k, F)$ are covered by \mathcal{S}' . \square

APPENDIX C

PROOF OF OBSERVATION 3

Proof. Assume otherwise. Then \mathcal{S}' covers at most $(1 - \frac{\epsilon}{100})n$ tasks of E_1 , each of which contributes 5 to the solution value, so the tasks of E_1 contribute at most $5n(1 - \frac{\epsilon}{100}) = 3m - \frac{\epsilon n}{20}$ value overall. From the above discussion, the tasks of E_2 contribute Ym , those of E_3 contribute Zm , and the single