This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing

CMAB-based Reverse Auction for Unknown Worker Recruitment in Mobile Crowdsensing

Mingjun Xiao, *Member, IEEE,* Baoyi An, Jing Wang, Guoju Gao,
Sheng Zhang, *Member, IEEE,* and Jie Wu, *Fellow, IEEE,*

Abstract—Mobile CrowdSensing (MCS), through which a requester can coordinate a crowd of workers to accomplish some data collection tasks, has been recognized as a promising paradigm for large-scale data acquisition in recent years. Many researches focus on the worker recruitment problem in MCS, but most of them either have the assumption that workers' qualities are known ahead of time or cannot ensure that workers report costs honestly. In this paper, we propose an incentive mechanism based on Combinatorial Multi-Armed Bandit and reverse Auction, called CMABA, to solve the multiple unknown workers recruitment problem in MCS. Our objective is to determine a recruiting strategy to maximize the total sensing quality under a limited budget, while ensuring truthfulness and individual rationality of sensing workers. We theoretically prove that our CMABA mechanism achieves truthfulness and individual rationality, and then analyze the regret of the mechanism. Based on CMABA, we ulteriorly propose an adaptive incentive mechanism, called ACMABA, to recruit workers via the alternative worker recruitment and quality update, which can achieve a higher total sensing quality and lower regret. Additionally, we also demonstrate significant performances of the CMABA and ACMABA mechanisms through extensive simulations on real-world data traces.

Index Terms—mobile crowdsensing, multi-armed bandit, reverse auction, worker recruitment.

1 Introduction

Mobile CrowdSensing (MCS) is a newly-emerging sensing paradigm for large-scale data collection, in which a crowd of mobile users are recruited to collect location-sensitive data with their smart devices when they visit some pre-designated places [2]. MCS systems have greater advantages than traditional systems for collecting economic value data distributed in a broad-scale area by leveraging users' mobility and diverse sensors embedded in smart devices. Many MCS mechanisms and systems have been designed recently to realize vast data collection in various scenarios such as environmental monitoring [3], smart transportation [4], spectrum sensing [5], etc.

A typical MCS system consists of a platform residing on the cloud, some data requesters, and a crowd of mobile users (a.k.a., workers). The platform can recruit workers to collect data according to the published data collection tasks from the requesters. However, the sensing qualities (qualities for short) vary for workers, even to execute the same task, because of the diverse performances of smart devices and the behaviors of mobile users. Thus, how to recruit workers with higher qualities is critical in the MCS systems. Most of the existing works focus on designing worker recruitment algorithms in the MCS systems [3], [6]–[8] under the circumstance that the quality information of workers is assumed to be known, which is unrealistic in practice. Moreover, it is unfeasible for a worker to evaluate its quality by itself in most cases. Since the workers' qualities are unknown a priori, some researches are working to investigate the quality unknown (unknown for short) worker recruitment problems in recent years [9]–[12].

Basically, a certain cost will be inevitably incurred by the smart devices usage and privacy leakage, etc. when collecting data, so the requesters need to provide adequate monetary rewards to incentivize workers to participate in the MCS systems. Moreover, rational workers might strategically report manipulated data collection costs to earn more rewards, because the cost information is generally private and sensitive. Hence, it is also vital to design suitable incentive mechanisms that can ensure workers to report costs honestly during the recruitment process. As we know, there are many incentive mechanisms that have been proposed to tackle the truthful cost report problem in MCS systems [13]–[15]. However, all of these incentive mechanisms are designed based on an assumption that the sensing qualities of workers are known a priori, while the qualities are actually unknown in practical worker recruitment scenarios.

In this paper, we consider the scenario that some quality unknown workers in MCS are recruited to execute the long-term data collection tasks (e.g., collect traffic flow, noise, air quality data, etc.) assigned by the requester under the budget constraint, where the costs of workers may be strategically reported. As the workers' qualities are unknown

- *M. Xiao, B. An, J. Wang are with the School of Computer Science and Technology / Suzhou Institute for Advanced Study, University of Science and Technology of China, Hefei, China.*
- *G. Gao is with the School of Computer Science and Technology, Soochow University, Suzhou, P. R. China.*
- *S. Zhang is with the State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China.*
- *J. Wu is with the Center for Networked Computing, Temple University, 1805 N. Broad Street, Philadelphia, PA 19122.*

This paper is an extended version of the conference paper [1] published in ICDCS 2020. This research was supported in part by the National Natural Science Foundation of China (NSFC) (Grant No. 61872330, 61572457, 61572336, 61632016, U1709217, 61379132), the Natural Science Foundation of Jiangsu Province in China (Grant No. BK20191194, BK20131174, BK2009150), and Anhui Initiative in Quantum Information Technologies (Grant No. AHY150300).

initially, the MCS platform will let workers execute some tasks for a few rounds and evaluate their collected data to learn their qualities, which is generally called *exploration*. On the other hand, the requester can obtain a higher valuation of data from the recruited workers with higher qualities. Because of this, the platform will recruit the best group of workers according to the learned knowledge of qualities so as to improve the total quality, which is known as *exploitation*. Therefore, we aim to design an incentive mechanism which can obtain the maximum total sensing quality within the limited budget, and ensure that workers report truthful costs to participate in MCS with high willingness.

Hence, there are three major challenges in the above incentive mechanism design for worker recruitment. The first challenge is how to determine an optimal trade-off between the exploration and exploitation phases under the shared budget to maximize the total sensing quality. Second, there are multiple periodical data collection tasks to be executed simultaneously, so the MCS platform needs to take account of the cost and quality factors in each round of worker recruitment. While considering the trade-off between exploration and exploitation, we also need to tackle the worker recruitment problem, especially for the exploitation phase. Third, to incentivize workers to participate in the data collection tasks and report their costs truthfully, some reasonable payment schemes needs to be designed for both of the exploration and exploitation phases.

To address the above challenges, we propose an incentive mechanism based on Combinatorial Multi-Armed Bandit and reverse Auction, called CMABA. In CMABA, the unknown worker recruitment problem is modeled as a Combinatorial Multi-Armed Bandit (CMAB) problem to tackle the trade-off between exploration and exploitation. Each worker is seen as an arm of CMAB, whose sensing quality is regarded as the corresponding reward, and the whole worker recruitment process is formulated as a combinatorial arm pulling process. To deal with the arm-pulling problem, we use the carefully defined Upper Confidence Bound (UCB) index to greedily select arms and compute the corresponding payments for winning arms. Based on CMABA, we also propose an adaptive incentive mechanism, called Adaptive CMABA (i.e., ACMABA), which conducts exploration and exploitation simultaneously in each arm-pulling round other than the necessary initial exploration. Overall, the major contributions are summarized as follows:

1) We propose two incentive mechanisms based on Combinatorial Multi-armed bandit and reverse Auction, namely CMABA and ACMABA. To the best of our knowledge, this is the first work that combines CMAB and reverse auction to solve the budget-feasible unknown worker recruitment problem in MCS systems.
2) We design a UCB-based greedy algorithm to recruit workers and compute payments for CMABA and ACMABA. Besides, we theoretically prove that they can achieve truthfulness, individual rationality and computational efficiency in each round of worker recruitment.
3) We analyze the regret bounds for CMABA and ACMABA respectively, and derive a near optimal budget allocation for exploration and exploitation in CMABA.
4) We conduct extensive simulations on real-world data traces to demonstrate the significant performances of the proposed two mechanisms.

The remainder of the paper is organized as follows. In Sec. 2, we introduce the system modeling and the problem formulation. The detailed design and theoretical analysis of CMABA is elaborated in Sec. 3 and Sec. 5. The ACMABA mechanism is introduced in Sec. 4. The simulations and evaluations are presented in Sec. 6. We review the related works in Sec. 7, and conclude the paper in Sec. 8.

2 System Model and Problem Formulation

2.1 System Overview

We consider a MCS system, composed of a platform, some requesters, and a crowd of unknown workers. A requester wants to recruit some workers to execute data collection tasks (e.g., collect traffic flow, noise, air quality data, etc.) in an urban area periodically within a limited budget B. The tasks and unknown workers are defined as follows:

Definition 1 (**Task, Weight, Round**)**.** The requester publishes M location-sensitive long-term data collection tasks via the platform, denoted by $\mathcal{M}{=}\{1,2,...,M\}$. Each task j ($\in\mathcal{M}$) is attached with a weight w_j to indicate its importance, where $\sum_{j\in\mathcal{M}} w_j{=}1$. Moreover, the data collection is divided into multiple rounds, denoted by $t\in\{1,2,\cdots\}$.

Definition 2 (**Unknown Worker, Sensing Quality, Cost**)**.** There are N unknown workers, denoted by $\mathcal{N}{=}\{1,2,...,N\}$. We let $q_{i,j}^t \in [0,1]$ denote the sensing quality of worker i ($\in\mathcal{N}$) completing task j ($\in\mathcal{M}$) in the t-th round. Each $q_{i,j}^t$ follows an unknown distribution with an unknown expectation q_i. Each worker i can perform a set of preferred tasks, denoted by $\mathcal{M}_i$, with a certain cost c_i for the whole tasks set. We assume that the cost for a worker to complete each task will not exceed a range $[c_{min}, c_{max}]$.

In Def. 2, the expected quality q_i is assumed to be fixed. But the sensing quality $q_{i,j}^t$ is determined by some worker-side factors, such as working ability, personal willingness, sensing context, and so on, which bring about the variance of actual sensing quality for different tasks [6], [16], [17]. Hence, for $\forall$ task $j'\neq j$, $q_{i,j'}^t$ may not be equal to $q_{i,j}^t$. For example, there are some workers taking photos in different places, where q_i depends on the lens, pixel, software, etc. of the camera embedded in its device, which is a fixed value. But the distance and angle of taking photo will make $q_{i,j}^t$ vary in different places even with the same device.

In addition, if the expected quality of each worker i varies in different tasks, we can use $|\mathcal{M}_i|$ virtual workers to replace this worker, where each virtual worker has an identical expected quality for each task. This can be equivalently regarded as the special case of $|\mathcal{M}_i|{=}1$. For example, if worker i executes the tasks in $\mathcal{M}_i{=}\{1,2,3\}$ with different expected qualities, the worker will be seen as three virtual workers i_1, i_2, i_3 to execute tasks $1,2,3$ with the expected qualities $q_{i_1}, q_{i_2}, q_{i_3}$, respectively. Each virtual worker can only execute one task.

The MCS system adopts the reverse auction to recruit unknown workers. The platform, the unknown workers and the requester are seen as the auctioneer, the workers and the buyer, respectively. The commodities for sale are the services of performing data collection tasks. First, workers submit the bids for their preferred tasks. Then, the platform

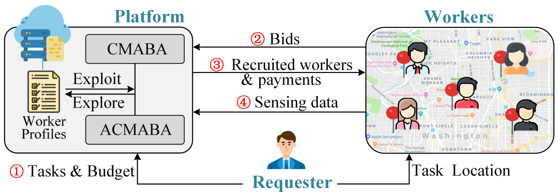


Fig. 1: The worker recruitment workflow of MCS.

determines the auction winners (i.e., makes the worker recruitment decision) and computes the corresponding payments, which are defined as follows:

Definition 3 (Bid). In each round t, each worker $i \in \mathcal{N}$ submits a task-bid pair $\beta_i{=}\langle \mathcal{M}_i, b_i \rangle$ to the platform, where b_i is the claimed cost for completing the tasks in $\mathcal{M}_i$. Since the true sensing cost c_i is private information, worker i will be a strategic worker who might misreport its cost to obtain higher utility. That is, the claimed cost b_i is not necessarily equal to the true cost c_i. Additionally, we denote all workers' bidding profiles as $\boldsymbol{\beta}{=}\{\beta_1, \beta_2, \cdots, \beta_N\}$.

Definition 4 (Payment). In each round t, the platform will determine a payment for the worker i who wins the auction, denoted as $p_{i,t}$. All workers' payment profiles are represented by $\boldsymbol{P}{=}\{\boldsymbol{P_1}, \boldsymbol{P_2}, \cdots, \boldsymbol{P_t}, \cdots\}$, in which $\boldsymbol{P_t}{=}\{p_{i,t} | i{\in}\mathcal{N}\}$.

Based on the reverse auction model, the detailed workflow of worker recruitment is illustrated in Fig.1. First, the requester publishes all sensing tasks and total budget on the platform. Second, all workers submit their task-bid pairs to the platform. Next, the worker recruitment process will be conducted round by round until the given budget is exhausted. In each round, the platform first conducts exploitation to determine the auction winners (i.e., selects the workers to be recruited) and calculates the corresponding payments. Then, the platform will inform the recruited workers to execute their preferred tasks. After completing the tasks, workers will return their sensing data to the requester via the platform. Since workers' sensing qualities are unknown, the platform needs to conduct exploration to learn these qualities according to the data. The qualities are recorded in the worker profiles and will be updated at the end of each round after receiving the sensed data.

2.2 CMAB Modeling and Problem Formulation

To determine auction winners and calculate the corresponding payments under the circumstance that all workers' qualities are unknown initially is the most critical issue for the platform over the whole worker recruitment process. On one hand, the platform needs to learn workers' quality values knowledge (i.e., so-called exploration) to obtain the better recruitment performance; on the other hand, the learned knowledge will be in turn utilized to make the best winners decision (i.e., so-called exploitation). This is actually an online learning and decision-making process. Multi-Armed Bandit (MAB) is a widely-used reinforcement learning model for online decision-making in uncertain environments [18]. A MAB model basically includes a slot machine with multiple arms, each of which is associated with a reward drawn from an unknown distribution. A player will pull some arms round by round according to a bandit policy, so as to maximize the cumulative reward. In

TABLE 1: Description of major notations

Variable	Description
$\mathcal{N}, \mathcal{M}$	the sets of workers and sensing tasks, respectively
N, M	the numbers of workers and sensing tasks, respectively
B	the total budget given by the requester
B'	the budget used for exploration
K	the numbers of workers recruited in each round
w_j	the weight of the j-th task
$\mathcal{M}_i$	the task set that worker i is willing to perform
c_i	the true cost of worker i completing $\mathcal{M}_i$
$q_{i,j}^t$	the quality of worker i completing task j in round t
q_i	the expected quality value of worker i
$\hat{q}_i^+$	the UCB index of worker i
$\hat{q}_{i,t}$	the sample mean of worker i's quality until round t
$\phi_{i,t}$	indicates whether worker i is recruited in round t
$p_{i,t}$	the payment to worker i in round t
M^+, M^-	the maximum and minimum values of $\lvert\mathcal{M}_i\rvert, \forall i \in \mathcal{N}$
RCR_i	the Revenue-Cost-Ratio of worker i

this paper, we model the unknown worker recruitment as a novel K-armed CMAB decision process, defined as follows:

Definition 5 (CMAB Modeling). The auction winner selection, which is unknown worker recruitment, is modeled as a K-armed CMAB. The platform is the player, each worker in $\mathcal{N}$ is an arm, and the quality of each recruited worker is seen as the reward of pulling the corresponding arm. In each round, the platform will pull K arms. Pulling the i-th arm indicates that worker i is a winner to be recruited.

Under the CMAB model, the worker recruitment problem is turned into a budget-feasible bandit policy determination problem so as to maximize the total expected revenue of the requester. The bandit policy and revenue are defined:

Definition 6 (Budget-feasible Bandit Policy). A bandit policy $\boldsymbol{\phi}$ is a sequence of maps: $\{\boldsymbol{\phi_1}, \boldsymbol{\phi_2}, \cdots, \boldsymbol{\phi_t}, \cdots\}$, where $\boldsymbol{\phi_t} = \{\phi_{1,t}, \phi_{2,t}, \cdots, \phi_{N,t}\}$. $\phi_{i,t} \in \{0, 1\}$ indicates whether worker i is selected in the t-th round. Moreover, the total cost of worker recruitment is no larger than the budget B.

Definition 7 (Revenue). Under the CMAB model, the total revenue refers to the total weighted sensing qualities of recruited workers, i.e., $R(\phi) = \sum_t \sum_{i=1}^{N} \sum_{j \in \mathcal{M}_i} w_j q_{i,j}^t \phi_{i,t}$. Hence, the total expected revenue is:

$$E\left[R(\phi)\right] = \sum_t \sum_{i=1}^{N} \sum_{j \in \mathcal{M}_i} w_j q_i \phi_{i,t} \tag{1}$$

Our objective is to maximize the total expected revenue under the budget constraint. Thus, the winner selection problem can be formulated as follows:

$$Maximize: \quad E\left[R(\phi)\right] \tag{2}$$

$$Subject\ to: \quad \sum_t \sum_{i=1}^{N} p_{i,t} \leq B \tag{3}$$

$$\sum_{i=1}^{N} \phi_{i,t} = K, \forall t \tag{4}$$

$$\phi_{i,t} \in \{0, 1\}, \forall i \in \mathcal{N}, \forall t \tag{5}$$

Here, Eq. (3) ensures that the total payment does not exceed the given budget. Eq. (4) indicates that K workers are recruited in each round. Eq. (5) implies that all worker recruitment decisions to be made are binary.

Furthermore, the platform needs to determine the auction winners' payments. To ensure that workers are willing to perform tasks and report their costs truthfully, the payment computation should meet truthfulness and individual rationality. Before this, we first define the concept of utility:

Definition 8 (Worker's Utility). The utility of worker $i \in \mathcal{N}$ in the t-th round is the payment minus cost:

$$u_{i,t} = (p_{i,t} - c_i)\,\phi_{i,t} \tag{6}$$

Here, $u_{i,t}$ and $p_{i,t}$ actually are the functions of bid b_i, i.e., $u_{i,t}$=$u_{i,t}(b_i)$, and $p_{i,t}$=$p_{i,t}(b_i)$.

Then, our mechanism satisfies the properties of truthfulness, individual rationality and computational efficiency.

Definition 9 (**Truthfulness**). Let b_i be an arbitrary bid of worker $i \in \mathcal{N}$ to complete tasks in $\mathcal{M}_i$, and $u_{i,t}(b_i)$ is the utility obtained by worker i in the t-th round. Then, if

$$u_{i,t}(b_i) \leq u_{i,t}(c_i), \forall i \in \mathcal{N}, \forall t \tag{7}$$

we say that the mechanism is truthful.

Definition 10 (**Individual Rationality**). For $\forall i \in \mathcal{N}$, if worker i's utility is non-negative,

$$u_{i,t} \geq 0, \forall i \in \mathcal{N}, \forall t \tag{8}$$

we say the mechanism is individual rationality.

Definition 11 (**Computational Efficiency**). A mechanism is computationally efficient if it generates the results and terminates in polynomial time.

For ease of reference, we list major notations in Table 1.

3 The CMABA Mechanism

In this section, we propose the incentive mechanism CMABA. We first introduce the basic idea of CMABA and then present the detailed algorithm, followed by an illustrative example to show how our CMABA mechanism works.

3.1 Basic Idea

Since the whole worker recruitment is an online learning and decision-making process, we separate it into the exploration and exploitation phases under the CMAB model. Accordingly, the budget is also divided into two parts. The first part of the budget is used solely to learn workers' qualities in the exploration phase, where we uniformly select the workers to execute tasks and pay each recruited worker the maximum cost he might incur. In the exploitation phase, all of the remaining budget is used to maximize the total expected revenue based on the acquired empirical knowledge, where we adopt a UCB-based greedy strategy to select workers and compute payments based on the second-price scheme to guarantee truthfulness. In what follows, the methods adopted in each phase are discussed separately.

3.1.1 Exploration Phase

In this phase, the platform tentatively recruits workers to perform sensing tasks, so as to learn their expected sensing qualities. Since the quality of each worker is unknown a priori, we need to treat each worker equally. We let the workers be explored in a round robin fashion. That is, workers $\{1, \cdots, K\}$ will be recruited in the first round and workers $\{K+1, \cdots, 2K\}$ will be recruited in the second round, and so on. We maintain two vectors $\boldsymbol{n_t} = \{n_{1,t}, \cdots, n_{N,t}\}$ and $\boldsymbol{\hat{q}_t} = \{\hat{q}_{1,t}, \cdots, \hat{q}_{N,t}\}$ as the empirical knowledge learned from the history. More specifically, $n_{i,t}$ is the number of times that i's quality is learned at the end of the t-th round and $\hat{q}_{i,t}$ is the sample mean of worker i's sensing quality by then. At the end of each round t, the platform observes the qualities of the recruited workers (i.e., $q_{i,j}^t$) and then updates the empirical knowledge. In our setting, once a worker i is recruited in a round, the corresponding quality would be learned $|\mathcal{M}_i|$ times. Thus, $n_{i,t}$ will be updated as follows:

$$n_{i,t} = \begin{cases} n_{i,t-1} + |\mathcal{M}_i|, & \phi_{i,t} = 1 \\ n_{i,t-1}, & \phi_{i,t} = 0 \end{cases} \tag{9}$$

The sample mean of worker i's quality is updated as

$$\hat{q}_{i,t} = \begin{cases} \frac{\hat{q}_{i,t-1} n_{i,t-1} + \sum_{j \in \mathcal{M}_i} q_{i,j}^t}{n_{i,t-1} + |\mathcal{M}_i|}, & \phi_{i,t} = 1 \\ \hat{q}_{i,t-1}, & \phi_{i,t} = 0 \end{cases} \tag{10}$$

The qualities learned in exploration will be used for worker recruitment in the exploitation phase. Here, we define UCB indexes for workers rather than using the sample means directly, denoted by $\boldsymbol{\hat{q}^+}$=$\{\hat{q}_1^+, \cdots, \hat{q}_N^+\}$, where we take the uncertainty of estimation into consideration. That is, the estimated sample mean may have a certain error, and the additive factor $\varepsilon_{i,t}$ makes the less selected worker have more chances to be selected. The UCB index $\hat{q}_i^+$ of worker i is

$$\hat{q}_i^+ = \hat{q}_{i,t} + \varepsilon_{i,t}, \quad \varepsilon_{i,t} = \sqrt{\frac{\delta \cdot \ln(\sum_{i' \in \mathcal{N}} n_{i',t})}{n_{i,t}}} \tag{11}$$

where δ is a positive hyper parameter that brings flexibility to our policy. This trick has also been used in [19], [20].

Since the exploration-separated algorithm relies significantly on the efficiency of the exploration phase, we need to carefully allocate the total budget among exploration and exploitation. We use B' to denote the budget dedicated to exploration. Let $M^+ = \max\{|\mathcal{M}_i| | i \in \mathcal{N}\}$, and M^-=$\min\{|\mathcal{M}_i| | i \in \mathcal{N}\}$. Then, B' is calculated by

$$B' = \left(\frac{1}{M^-}\right)^{\frac{1}{3}} \left(\delta \cdot N M^+ c_{max} \ln\left(\frac{M^+ B}{M^- c_{max}}\right)\right)^{\frac{1}{3}} B^{\frac{2}{3}} \tag{12}$$

Here, B' is calculated to minimize the regret of the algorithm, which is detailedly presented in Sec. 5. For each recruited worker i in the exploration phase, we set the maximum cost that he might incur as the payment, i.e. $|\mathcal{M}_i| c_{max}$, to guarantee truthfulness and individual rationality. The exploration process terminates until B' is exhausted.

3.1.2 Exploitation Phase

In the exploitation phase, we determine the winning workers and the corresponding payments based on the UCB indexes learned in the exploration phase. We assume that all workers report their costs truthfully, i.e., b_i=c_i, which is to be proved reasonable in Sec. 5. Since the worker recruitment problem in our paper can be regarded as a series of special 0-1 knapsack problems, we adopt a UCB-based greedy strategy to select workers. For each worker i, we define the Revenue-Cost-Ratio (RCR henceforth) as $(\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+)/b_i$. The idea for recruiting workers is to select the workers with the highest RCRs in each round. Thus, we first calculate the RCR for each worker and then sort the workers in $\mathcal{N}$ into $\boldsymbol{S}$=$(s_1, s_2, \cdots, s_N)$ such that $RCR_{s_1} \geq \cdots \geq RCR_{s_N}$. Then, we greedily select the best K workers into the winning workers set $\boldsymbol{S'}$ and compute the payments for them.

To guarantee truthfulness, each winning worker should be paid the critical payment according to [21]. The critical payment is equal to the highest bid that still makes the bid win. More specifically, the critical payment of a winning worker i should be calculated based on the bid of the $(K+1)$-th worker in $\boldsymbol{S}$, i.e., $\frac{\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+}{\sum_{j \in \mathcal{M}_{s_{K+1}}} w_j \hat{q}_{s_{K+1}}^+} b_{s_{K+1}}$. When the critical payment for worker i is larger than $|\mathcal{M}_i| c_{max}$, we set the payment as $|\mathcal{M}_i| c_{max}$. Hence, $\forall i \in \mathcal{N}$, the payment of worker i in the exploitation phase is calculated as:

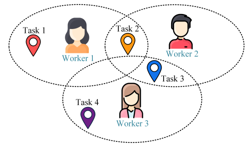
$$p_i = \min\left\{\frac{\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+}{\sum_{j \in \mathcal{M}_{s_{K+1}}} w_j \hat{q}_{s_{K+1}}^+} b_{s_{K+1}}, |\mathcal{M}_i| c_{max}\right\} \tag{13}$$

The exploitation terminates until the budget exhausts.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing

Algorithm 1: The CMABA Mechanism

Input: $\mathcal{N}, \mathcal{M}, B, K, \boldsymbol{\beta}, c_{max}, \{w_j | j \in \mathcal{M}\}, \delta$
Output: $\boldsymbol{\phi}, \boldsymbol{P}$
1 Initialize $t = 0$; $\phi_{i,t} = 0, p_{i,t} = 0, \forall i \in \mathcal{N}, \forall t$;
2 $B' = (\frac{1}{M^-})^{\frac{1}{3}} (\delta \cdot NM^+ c_{max} \ln(\frac{M^+ B}{M^- c_{max}}))^{\frac{1}{3}} B^{\frac{2}{3}}$;
$B_t = B'$; // Exploration phase
3 **while** *true* **do**
4 $t = t + 1; \mathcal{K} = \emptyset$;
5 **for** $j = 1 : K$ **do**
6 $i = ([(t-1)K + j - 1] mod N) + 1; \mathcal{K} = \mathcal{K} \cup \{i\}$;
7 **if** $\sum_{i \in \mathcal{K}} |\mathcal{M}_i| c_{max} \le B_{t-1}$ **then**
8 **foreach** $i \in \mathcal{K}$ **do** $\phi_{i,t} = 1$; $p_{i,t} = |\mathcal{M}_i| c_{max}$;
9 **foreach** $i \in \mathcal{N}$ **do** Update $n_{i,t}$, $\hat{q}_{i,t}$, $\hat{q}_i^+$;
10 Update B_t;
11 **else** break;
12 $B_t = B - B' + B_{t-1}$; // Exploitation phase
13 **foreach** $i \in \mathcal{N}$ **do** Calculate the RCR: $\frac{\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+}{b_i}$;
14 Sort the workers $\mathcal{N}$ into $\boldsymbol{S}$: $RCR_{s_1} \ge \cdots \ge RCR_{s_N}$;
15 Select the best K workers: $\boldsymbol{S'} = \{s_1, s_2, \cdots, s_K\}$;
16 **foreach** $s_i \in \boldsymbol{S'}$ **do**
17 $p_{s_i} = \min\{\frac{\sum_{j \in \mathcal{M}_{s_i}} w_j \hat{q}_{s_i}^+}{\sum_{j \in \mathcal{M}_{s_{K+1}}} w_j \hat{q}_{s_{K+1}}^+} b_{s_{K+1}}, |\mathcal{M}_{s_i}| c_{max}\}$;
18 **while** $\sum_{s_i \in \boldsymbol{S'}} p_{s_i} \le B_t$ **do**
19 **foreach** $s_i \in \boldsymbol{S'}$ **do** $\phi_{s_i,t} = 1$; $p_{s_i,t} = p_{s_i}$;
20 $t = t + 1$; $B_t = B_{t-1} - \sum_{s_i \in \boldsymbol{S'}} p_{s_i,t-1}$;
21 **return** $(\boldsymbol{\phi}, \boldsymbol{P})$;



(a) Locations of tasks and workers

	Worker 1	Worker 2	Worker 3
Preferred tasks	{1,2}	{2,3}	{3,4}
Claimed costs	0.5	1	1.2
Expected qualities	0.6	0.7	0.8

(b) Workers' information

Fig. 2: Information about workers and tasks

	Round 1		Round 2		Round 3		Round 4		Round 5...
$q_{i,j}^t$	Worker 1	Worker 2	Worker 3	Worker 1	Worker 2	Worker 3	Worker 3	Worker 1	...
Task 1	0.7			0.8				0.75	...
Task 2	0.4	0.48		0.5	0.62			0.6	...
Task 3		0.7	0.9		0.8	0.8	0.68		...
Task 4			0.64			0.58	0.83		...

Fig. 3: The actual qualities in different rounds

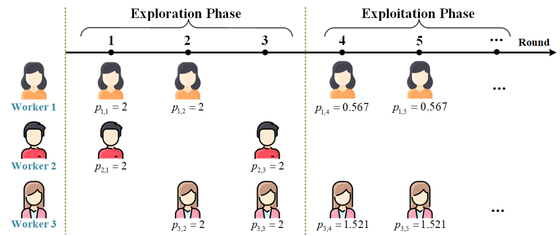


Fig. 4: The whole worker recruitment process

3.2 The Detailed Algorithm

As shown in Algorithm 1, we first initialize all $\phi_{i,t}$=0, $p_{i,t}$=0 (Step 1). Then in the exploration phase, we calculate the exploration budget B' according to Eq. (12) (Step 2), and tentatively recruit workers to complete tasks (Steps 3-8). At the end of each round t, we observe each recruited worker i's quality values $\{q_{i,j}^t | j \in \mathcal{M}_i\}$, and update $\boldsymbol{n_t}, \boldsymbol{\hat{q}_t}, \boldsymbol{\hat{q}^+}$ according to Eqs. (9)-(11), respectively (Step 9). The exploration terminates until B' cannot afford the next round.

In the exploitation phase, we leverage the quality information learned in the exploration phase to select workers. Except for the dedicated exploitation budget, we recycle the remaining budget in the exploration phase (Step 12). In Steps 13-15, we first calculate the RCR for each worker and sort the workers in a non-increasing order of their RCRs. Then, we greedily select the best K workers into a winning set $\boldsymbol{S'}$ and compute the corresponding payments according to Eq. (13) (Steps 16-18). Next, the algorithm continuously checks if the total payment for the winning workers is smaller than the current leftover budget (Steps 19-22). If so, it employs the workers in the winning set to perform tasks for the current round and subtracts their payments from the current budget. Otherwise, the algorithm stops and returns the worker recruitment result as well as the payment profile.

3.3 An illustrative Example

For a better understanding, we provide a simple example to illustrate the worker recruitment process of the CMABA mechanism. We consider the scenario where there are three sensing workers $\mathcal{N}=\{1,2,3\}$ and four sensing tasks $\mathcal{M}=\{1,2,3,4\}$, as depicted in Fig.2(a). Each worker is associated with a service coverage, indicating that the worker can only conduct tasks within this coverage, which is shown as a circle in Fig.2(a). In this example, we assume that a worker is willing to perform all tasks within his coverage. Thus, we have $\mathcal{M}_1=\{1,2\}, \mathcal{M}_2=\{2,3\}, \mathcal{M}_3=\{3,4\}$. Here, we set the weight of each task as w_1=0.1, w_2=0.2, w_3=0.3 and w_4=0.4. In addition, we set the monetary budget as 50$ and let K=2, $\delta=\frac{1}{8}$. The claimed costs and the expected qualities (which are unknown a priori) are shown in Fig.2(b). We assume that the quality of each worker follows the uniform distribution in [0,1], and the cost for a worker completing each task would not exceed $[0.1, 1]$, i.e., c_{min}=0.1 and c_{max}=1.

At the beginning, we do not have any information about workers' qualities. Thus, we first calculate the exploration budget as $(\frac{1}{2})^{\frac{1}{3}} \cdot (\frac{1}{8} \cdot 3 \cdot 2 \cdot 1 \cdot \ln(50))^{\frac{1}{3}} \cdot (50)^{\frac{2}{3}} = 15\$$ according to Eq. (12) and then learn the quality information. According to Algorithm 1, the workers will be recruited in the order of $\langle 1,2 \rangle, \langle 3,1 \rangle, \langle 2,3 \rangle$ in the exploration phase. Since there are 2 tasks in each task set, every time a worker is recruited, he will be paid 2$ and the corresponding quality will be learned twice. The actual qualities of each recruited worker in different rounds are presented in Fig.3. After the first three rounds, the remaining exploration budget is 3$ and is not sufficient to support the next exploration round. The sample quality means for three workers after the third round are: $\hat{q}_{1,3} = (0.7+0.4+0.8+0.5)/4 = 0.6$, $\hat{q}_{2,3} = (0.48+0.7+0.62+0.8)/4 = 0.65$, $\hat{q}_{3,3} = (0.9+0.64+0.8+0.58)/4 = 0.73$. Accordingly, the UCB index for each worker is calculated as: $\hat{q}_1^+ = 0.87, \hat{q}_2^+ = 0.92, \hat{q}_3^+ = 1$. Next, the exploitation phase begins and the residual budget is $35+3 = 38\$$. We first calculate the RCR for each worker: $RCR_1 = (0.1+0.2) \cdot 0.87/0.5 = 0.522, RCR_2 = (0.2+0.3) \cdot 0.92/1 = 0.46, RCR_3 = (0.3+0.4) \cdot 1/1.2 = 0.583$. Since $RCR_3 > RCR_1 > RCR_2$, worker 3 and worker 1 are the winners and will always be recruited until the remaining budget runs out. Their payments are calculated according to Eq. (13): $p_{3,4} = min\{\frac{(0.3+0.4) \cdot 1}{(0.2+0.3) \cdot 0.92} \cdot 1, 2\} = 1.521, p_{1,4} = min\{\frac{(0.1+0.2) \cdot 0.87}{(0.2+0.3) \cdot 0.92} \cdot 1, 2\} = 0.567$. Since worker 2 is not re-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing

Algorithm 2: The ACMABA Mechanism

Input: $\mathcal{N}, \mathcal{M}, B, K, \boldsymbol{\beta}, c_{max}, \{w_j | j \in \mathcal{M}, \delta\}$
Output: $\boldsymbol{\phi}, \boldsymbol{P}$
1 Initialize $t = 0$; $\phi_{i,t} = 0, p_{i,t} = 0, \forall i \in \mathcal{N}, \forall t$;
2 $B_t = B$; // Initial exploration phase
3 **while** $t \leq \frac{N}{K}$ **do**
4 $\quad t = t + 1; \mathcal{K} = \emptyset$;
5 $\quad$ **for** $j = 1 : K$ **do**
6 $\quad\quad i = ([(t-1)K + j - 1] mod N) + 1; \mathcal{K} = \mathcal{K} \cup \{i\}$;
7 $\quad$ **if** $\sum_{i \in \mathcal{K}} |\mathcal{M}_i| c_{max} \leq B_{t-1}$ **then**
8 $\quad\quad$ **foreach** $i \in \mathcal{K}$ **do** $\phi_{i,t} = 1$; $p_{i,t} = |\mathcal{M}_i| c_{max}$;
9 $\quad\quad$ **foreach** $i \in \mathcal{N}$ **do** Update $n_{i,t}$, $\hat{q}_{i,t}$, $\hat{q}_i^+$;
10 $\quad\quad$ Update $B_t = B_{t-1} - \sum_{i \in \mathcal{K}} p_{i,t}$;
11 $\quad$ **else** $\perp$;
12 **while** t++ **do**
13 $\quad$ **foreach** $i \in \mathcal{N}$ **do** Calculate the RCR: $\frac{\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+}{b_i}$;
14 $\quad$ Sort the workers $\mathcal{N}$ into $\boldsymbol{S}$: $RCR_{s_1} \geq \cdots \geq RCR_{s_N}$;
15 $\quad$ Select the best K workers: $\boldsymbol{S'} = \{s_1, s_2, \cdots, s_K\}$;
16 $\quad$ **foreach** $s_i \in \boldsymbol{S'}$ **do**
17 $\quad\quad p_{s_i} = \min\{\frac{\sum_{j \in \mathcal{M}_{s_i}} w_j \hat{q}_{s_i}^+}{\sum_{j \in \mathcal{M}_{s_{K+1}}} w_j \hat{q}_{s_{K+1}}^+} b_{s_{K+1}}, |\mathcal{M}_{s_i}| c_{max}\}$;
18 $\quad$ **if** $\sum_{s_i \in \boldsymbol{S'}} p_{s_i} \leq B_{t-1}$ **then**
19 $\quad\quad$ **foreach** $s_i \in \boldsymbol{S'}$ **do** $\phi_{s_i,t} = 1$; $p_{s_i,t} = p_{s_i}$;
20 $\quad\quad$ Update $B_t = B_{t-1} - \sum_{s_i \in \boldsymbol{S'}} p_{s_i,t}$;
21 $\quad\quad$ **foreach** $i \in \mathcal{N}$ **do** Update $n_{s_i,t}$, $\hat{q}_{s_i,t}$, $\hat{q}_{s_i}^+$;
22 $\quad$ **else** break;
23 **return** $(\boldsymbol{\phi}, \boldsymbol{P})$;

cruited, the corresponding payment is $p_2 = 0$. Hence, the recruitment order is $\{\langle 1, 2\rangle, \langle 3, 1\rangle, \langle 2, 3\rangle, \langle 3, 1\rangle, \cdots, \langle 3, 1\rangle\}$ and the whole recruitment process is shown in Fig.4.

Suppose the quality information is known in advance. We can see that workers 1 and 3 are actually the best two workers. Moreover, we can observe that the payment to each recruited worker in each round is no less than its true sensing cost, which satisfies individual rationality.

4 Extension

The CMABA mechanism fits better in the scenario that the workers' qualities are stable, implying that the variance of workers' qualities is small. To tackle another possible scenario that the variance of workers' qualities is relatively high, we extend the CMABA mechanism to obtain better performances in total revenue and regret with a little loss of truthfulness, where adaptive quality is taken into account in the exploitation phase. Based on the issue, we propose the **A**daptive **C**ombinatorial **M**ulti-**A**rmed **B**andit and reverse **A**uction (ACMABA) mechanism for the extended problem.

4.1 Basic Idea

Under the circumstance that variance of different workers' qualities is relatively high, the workers' qualities need to be learned more times so as to make the estimated quality more precise. We consider adding the quality learning across the auction rounds described in Sec. 2, which means that we need to design an adaptive and incentive mechanism to iteratively conduct the online learning and decision-making in the whole worker recruitment. Then, we propose the ACMABA mechanism based on CMABA, which alternates exploration and exploitation simultaneously in each round.

In the initial exploration phase, given the budget B, the platform needs to execute an initial quality learning for all workers $\mathcal{N}$ until all workers have been learned at least once due to the unknown workers' qualities. Compared with the exploration phase in Sec. 3.1.1, we similarly recruit K workers in each round and update each worker i's quality based on the learned $n_{i,t}$ and $\hat{q}_{i,t}$ according to Eqs. (9) and (10) at the end of the t-th round. Until each worker has been recruited and the corresponding quality has been learned at least once, the platform will exploit the learned UCB indexes computed by Eq. (11) to execute a reverse auction in the next round to determine K winning workers based on each worker i's **RCR** (i.e., $\frac{\sum_{j \in \mathcal{M}_i} w_j \hat{q}_i^+}{b_i}$), and compute the corresponding payment for each winning worker i based on the bid of the $(K+1)$-th worker and Eq. (13). After the auction in the exploitation phase, the platform observes the qualities of the recruited workers and updates the workers' qualities and UCB indexes based on the observed qualities and Eqs. (9), (10) and (11) in the exploration phase. In the following rounds, the platform will similarly in turn conduct the reverse auction (i.e., exploitation) and the quality learning (i.e., exploration) until the total budget is exhausted.

4.2 The Detailed Algorithm

According to the above solution, the proposed ACMABA mechanism is depicted in Algorithm 2. At the beginning, we initialize all $\phi_{i,t} = 0, p_{i,t} = 0$ (Step 1). Then, the initial exploration phase begins. We first tentatively recruit each worker to complete tasks in order to learn and estimate their quality values (Steps 3-8). At the end of each round t, we observe $\{q_{i,j}^t | j \in \mathcal{M}_i\}$ for each recruited worker i, and then update $\boldsymbol{n_t}, \boldsymbol{\hat{q}_t}, \boldsymbol{\hat{q}}^+$ according to Eqs. (9)-(11), respectively (Step 10). The initial exploration phase terminates until each worker has been selected at least once.

We have preliminarily learned the quality information of each worker in the initial exploration phase. Then in the next round, we can use the quality information to select workers and compute payments (i.e., exploitation). In Steps 13-14, we first calculate the RCR for each worker and then sort the workers in a non-increasing order of their RCRs. Then, we greedily select the best K workers into a winning set $\boldsymbol{S'}$ and determine the corresponding payments according to (13) (Steps 15-17). Next, the algorithm checks if the total payment for the winning workers is smaller than the current leftover budget (Steps 18). If so, it employs the workers in the winning set to perform sensing tasks for the current round and subtracts their payments from the current budget (Steps 19-20). Besides, the algorithm will also update $\boldsymbol{n_t}, \boldsymbol{\hat{q}_t}, \boldsymbol{\hat{q}}^+$ simultaneously based on the observed qualities in the same current round (Step 21, i.e., exploration). The algorithm alternates between the auction (i.e., exploitation) and the quality update (i.e., exploration) in each of the subsequent rounds until the remaining budget cannot afford the payments. So far, the algorithm returns the worker recruitment result and the payment profile.

4.3 An illustrative Example

For better comparison, the basic parameters setting in the walkthrough example, which illustrates the worker recruit-

	Round 1		Round 2		Round 3		Round 4		Round 5...
$q_{i,j}^t$	Worker 1	Worker 2	Worker 3	Worker 1	Worker 3	Worker 1	Worker 3	Worker 2	...
Task 1	0.7			0.8		0.58			...
Task 2	0.4	0.48		0.5		0.6		0.7	...
Task 3		0.7	0.9		0.7		0.68	0.6	...
Task 4			0.64		0.58		0.62		...

Fig. 5: The actual qualities in different rounds

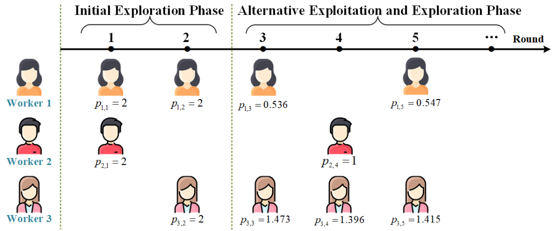


Fig. 6: The whole worker recruitment process

ment process of the ACMABA mechanism, is the same as the setting in the CMABA mechanism, as shown in Figs. 2(a) and 2(b). The actual qualities of each recruited worker in different rounds change a little as presented in Fig. 5.

At the beginning, we do not have any information about workers' qualities. Thus, we first conduct an initial exploration to learn the quality information. According to Algorithm 2, the workers will be recruited in the order of $\langle 1,2\rangle, \langle 3,1\rangle$ in the initial exploration phase. Since there are 2 tasks in each task set, every time a worker is recruited, he will be paid 2$ and the corresponding quality will be learned twice. Until the quality information of each worker has been learned at least once, the initial exploration stops at the end of the second round. The sample quality means for three workers after the second round are: $\hat{q}_{1,2}=(0.7+0.4+0.8+0.5)/4=0.6$, $\hat{q}_{2,2}=(0.48+0.7)/2=0.59$, $\hat{q}_{3,2}=(0.9+0.64)/2=0.77$. Accordingly, the UCB index for each worker is calculated as: $\hat{q}_1^+=0.85$, $\hat{q}_2^+=0.95$, $\hat{q}_3^+=1$.

The ACMABA mechanism can enter the adaptive worker recruitment which is dominated by the auction (i.e., exploitation) alternated with quality learning (i.e., exploration). In round 3, the exploitation phase begins and the residual budget is $50-8=42$$. We first calculate the RCR for each worker: $RCR_1=(0.1+0.2)\cdot 0.85/0.5=0.51$, $RCR_2=(0.2+0.3)\cdot 0.95/1=0.475$, $RCR_3=(0.3+0.4)\cdot 1/1.2=0.58$. Since $RCR_3>RCR_1>RCR_2$, worker 3 and worker 1 are the winners at the end of round 3. Their payments are calculated according to (13): $p_{3,3}=min\{\frac{(0.3+0.4)\cdot 1}{(0.2+0.3)\cdot 0.95}\cdot 1,2\}=1.473$, $p_{1,3}=min\{\frac{(0.1+0.2)\cdot 0.85}{(0.2+0.3)\cdot 0.95}\cdot 1,2\}=0.536$. Since worker 2 is not recruited, the corresponding payment is $p_{2,3}=0$. Next, the ACMABA conducts exploration to update qualities, so that $\hat{q}_{1,3}=(0.6\cdot 4+0.58+0.6)/6=0.59$, $\hat{q}_{2,3}=\hat{q}_{2,2}=0.59$, $\hat{q}_{3,3}=(0.77\cdot 2+0.7+0.58)/4=0.7$. Then, the UCB index for each worker is: $\hat{q}_1^+=0.8175$, $\hat{q}_2^+=0.9841$, $\hat{q}_3^+=0.9787$.

In round 4, we can similarly calculate the RCR for each worker: $RCR_1=0.4905, RCR_2=0.4920, RCR_3=0.5709$. Since $RCR_3>RCR_2>RCR_1$, worker 3 and worker 2 are the winners at the end of round 4. Their payments are calculated according to (13): $p_{3,4}=1.396, p_{2,4}=1$, where $p_{1,4}=0$. Then, the qualities and UCB indexs can be updated in the same way, which are $\hat{q}_{1,4}=0.59$, $\hat{q}_{2,4}=0.62$, $\hat{q}_{3,4}=0.68$ and $\hat{q}_1^+=0.83$, $\hat{q}_2^+=0.91$, $\hat{q}_3^+=0.92$. In round 5, we can easily obtain: $RCR_1=0.498, RCR_2=0.455, RCR_3=0.536$. So, worker 3 and worker 1 are the winners at the end of round 5, where their payments are: $p_{1,5}=0.547$, $p_{2,5}=0$, $p_{3,5}=1.415$. So far, the recruitment order at the end of the 4-th round is $\{\langle 1,2\rangle, \langle 3,1\rangle, \langle 3,1\rangle, \langle 3,2\rangle, \langle 3,1\rangle, \cdots\}$ and the whole recruitment process is shown in Fig.6. Compared with the CMABA mechanism, we can notice that the recruited workers are continuously changing as the qualities update.

5 Performance Analysis

In this section, we present the theoretical analysis for CMABA and ACMABA. We first analyze the worst regret bounds of these two algorithms respectively, and then prove that they satisfy the properties of truthfulness, individual rationality and computational efficiency.

5.1 Regret Analysis

Regret is the difference between the revenue obtained by the proposed algorithms and the optimal solution. Assume that workers' qualities are known in advance. The optimal total revenue cannot be feasibly achieved. However, the platform can achieve an approximately optimal solution by recruiting the workers with the highest RCRs in each round, which is denoted by ϕ^*. Here, $R(\phi^*)\geq\alpha\cdot R(\phi_{opt})$, where ϕ_{opt} is the optimal recruiting policy that can obtain the maximum revenue and $0<\alpha\leq 1$. In this case, it is no longer fair to compare the performance of the proposed algorithms against the optimal solution. Therefore, we introduce the concept of α-approximation regret [22], [23] for our algorithms:

$$R_\alpha=\alpha\cdot R(\phi_{opt})-E[R(\phi)]\leq R(\phi^*)-E[R(\phi)] \quad (14)$$

According to [24], [25], we have $\alpha\geq\frac{1}{2}$ in our case.

When the qualities are known, we denote the worker set as $\boldsymbol{V}=\{v_1,\cdots,v_N\}$ such that $RCR'_{v_1}\geq\cdots\geq RCR'_{v_N}$. Here, the RCR' of worker $i\in\boldsymbol{V}$ is calculated as $(\sum_{j\in\mathcal{M}_i} w_j q_i)/b_i$. The payment of i is $p'_i=(\sum_{j\in\mathcal{M}_i} w_j q_i)/RCR'_{v_{K+1}}$. The α-optimal algorithm will recruit the best K workers $\{v_1,\cdots,v_K\}$ in each round until the budget exhausts. Thus, the total revenue obtained by the α-optimal algorithm is:

$$R(\phi^*)=\frac{B}{\sum_{i=v_1}^{v_K}p'_i}\cdot(\sum_{i=v_1}^{v_K}\sum_{j\in\mathcal{M}_i}w_j q_i)=B\cdot RCR'_{v_{K+1}} \quad (15)$$

Recall that $M^+=\max\{|\mathcal{M}_i|\}$, $M^-=\min\{|\mathcal{M}_i|\}$. Let $w^+=\max\{w_j\}$, $w^-=\min\{w_j\}$. Before the detailed regret analysis, we introduce some auxiliary lemmas.

Lemma 12 (Chernoff-Hoeffding bound)**.** *Suppose that* $\{X_1,X_2,\cdots,X_n\}$ *are* n *random variables with common range* $[0,1]$, *satisfying* $E[X_t|X_1,\cdots,X_{t-1}]=\mu$ *for* $\forall t\in[1,n]$. *Let* $S_n=X_1+\cdots+X_n$. *Then* $\forall a\geq 0$

$$P[S_n\geq n\mu+a]\leq e^{-2a^2/n}, P[S_n\leq n\mu-a]\leq e^{-2a^2/n}. \quad (16)$$

5.2 Regret of CMABA

Let n_i be the number of learned times, and $\hat{q}_i$ be the sample mean of worker i's quality after exploration, respectively. We denote the maximum and minimum exploration rounds as r^+ and r^-, respectively. Actually, $r^-=\lfloor\frac{B'}{KM^+c_{max}}\rfloor$, $r^+=\lfloor\frac{B'}{KM^-c_{max}}\rfloor$. Then, we have $\lfloor\frac{r^-K}{N}\rfloor|\mathcal{M}_i|\leq n_i\leq\lceil\frac{r^+K}{N}\rceil|\mathcal{M}_i|$.

Lemma 13. *After exploration,* $q_i<\hat{q}_i^+$ *holds for each* $i\in\mathcal{N}$ *with the probability of at least* $1-N(NM^-(\frac{r^-K}{N}-1))^{-2\delta}$.

Proof. According to Lemma 12, we can derive

$$\begin{aligned}P(\hat{q}_i^+\leq q_i)&=P(\hat{q}_i+\sqrt{\frac{\delta\cdot\ln(\sum_{i'\in\mathcal{N}}n_{i'})}{n_i}}\leq q_i)\\&\leq e^{-2n_i(\sqrt{\frac{\delta\cdot\ln(\sum_{i'\in\mathcal{N}}n_{i'})}{n_i}})^2}\leq e^{-2\delta\cdot\ln(N\lfloor\frac{r^-K}{N}\rfloor M^-)}\\&\leq(NM^-(\frac{r^-K}{N}-1))^{-2\delta}\end{aligned} \quad (17)$$

Therefore, $\forall i \in \mathcal{N}$, $q_i < \hat{q}_i^+$ with the probability of at least $1-(NM^-(\frac{r^-K}{N}-1))^{-2\delta}$. So we can obtain that

$$\begin{aligned} &P(\cap_{i\in\mathcal{N}}(q_i<\hat{q}_i^+))=1-P(\cup_{i\in\mathcal{N}}(q_i\geq\hat{q}_i^+)) \\ &\geq 1-\sum_{i\in\mathcal{N}}P(q_i\geq\hat{q}_i^+) \ \geq 1-N(NM^-(\frac{r^-K}{N}-1))^{-2\delta}, \quad (18) \end{aligned}$$

which proves the lemma. □

Lemma 14. *With the probability of at least* $1-N(NM^-(\frac{r^-K}{N}-1))^{-2\delta}$, $\forall l\in[1,N]$, $RCR'_{v_l}\leq RCR_{s_l}$.

Proof. According to Lemma 13, we can derive that for each worker $i \in \mathcal{N}$, $\frac{\sum_{j\in\mathcal{M}_i} w_j q_i}{b_i} < \frac{\sum_{j\in\mathcal{M}_i} w_j \hat{q}_i^+}{b_i}$ holds with the probability of at least $1-N(NM^-(\frac{r^-K}{N}-1))^{-2\delta}$. Then we prove Lemma 14 as follows. There are two cases:

Case 1: s_l and v_l are the same worker. The lemma holds.

Case 2: s_l and v_l are different workers. Assume for purpose of contradiction that the lemma does not hold, i.e., $\frac{\sum_{j\in\mathcal{M}_{v_l}} w_j q_{v_l}}{b_{v_l}} > \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \hat{q}_{s_l}^+}{b_{s_l}}$. Then we have $\frac{\sum_{j\in\mathcal{M}_{v_1}} w_j \hat{q}_{v_1}^+}{b_{v_1}} > \frac{\sum_{j\in\mathcal{M}_{v_1}} w_j q_{v_1}}{b_{v_1}} \geq \cdots \geq \frac{\sum_{j\in\mathcal{M}_{v_l}} w_j q_{v_l}}{b_{v_l}} > \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \hat{q}_{s_l}^+}{b_{s_l}}$.We can see that $\frac{\sum_{j\in\mathcal{M}_{v_1}} w_j \hat{q}_{v_1}^+}{b_{v_1}} > \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \hat{q}_{s_l}^+}{b_{s_l}}$. For the same reason, we can also derive $\frac{\sum_{j\in\mathcal{M}_{v_2}} w_j \hat{q}_{v_2}^+}{b_{v_2}} > \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \hat{q}_{s_l}^+}{b_{s_l}}, \dots, \frac{\sum_{j\in\mathcal{M}_{v_l}} w_j \hat{q}_{v_l}^+}{b_{v_l}} > \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \hat{q}_{s_l}^+}{b_{s_l}}$.Therefore, there are already l workers whose RCRs are larger than s_l's RCR, and s_l will be ranked behind the l-th worker, which contradicts with the definition that s_l is the l-th worker in $\boldsymbol{S}$. Thus, the lemma holds. □

Theorem 15. *The worst α-approximation regret of the proposed CMABA algorithm is bounded as $O(B^{\frac{2}{3}}(NM^4\ln(BM))^{\frac{1}{3}})$ with the probability of at least $1-N(NM^-(\frac{r^-K}{N}-1))^{-2\delta}$.*

Proof. The expected regret of CMABA can be bounded as

$$\begin{aligned} R(\phi^*)-E[R(\phi)] &\leq B\cdot RCR'_{v_{K+1}}-\Big(\lfloor\tfrac{r^-K}{N}\rfloor(\textstyle\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{M}_i}w_j q_i) \\ &\quad +\lfloor\tfrac{B-B'}{\sum_{i\in S'}p_i}\rfloor\cdot(\textstyle\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j q_i)\Big) \\ &< B'\cdot RCR'_{v_{K+1}}+(B-B')\cdot\Big(RCR'_{v_{K+1}}-\tfrac{\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j q_i}{\sum_{i\in S'}p_i}\Big) \\ &= B'\cdot RCR'_{v_{K+1}}+\tfrac{B-B'}{\sum_{i\in S'}p_i} \\ &\quad \cdot\Big(\tfrac{RCR'_{v_{K+1}}}{RCR_{s_{K+1}}}(\textstyle\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+)-\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j q_i\Big) \\ &\leq B'\cdot RCR'_{v_{K+1}}+\tfrac{B-B'}{\sum_{i\in S'}p_i}(\textstyle\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j(\hat{q}_i^+-q_i)) \end{aligned}$$

The last inequality above is based on Lemma 14. Then, according to the individual rationality proved in Theorem 20, we have $p_i \geq c_i, \forall i \in \mathcal{N}$. Hence,

$$\begin{aligned} R_\alpha &< B'\cdot RCR'_{v_{K+1}}+\tfrac{B-B'}{\sum_{i\in S'}c_i}(\textstyle\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}w_j(\hat{q}_i^+-q_i)) \\ &< B'\cdot RCR'_{v_{K+1}}+\tfrac{B-B'}{\sum_{i\in S'}c_i}\cdot(\textstyle\sum_{i\in S'}\sum_{j\in\mathcal{M}_i}2w_j\sqrt{\tfrac{\delta\ln(\sum_{i'\in\mathcal{N}}n_{i'})}{n_i}}) \end{aligned}$$

(according to the Chernoff-Hoeffding bound in Lemma 12)

$$\leq \frac{B'M^+w^+}{M^-c_{min}}+\frac{B-B'}{KM^-c_{min}}\cdot 2KM^+w^+\sqrt{\frac{\delta\ln(N\lceil\frac{r^+K}{N}\rceil M^+)}{\lfloor r^-K/N\rfloor M^-}}$$

Substituting r^- with $\lfloor\frac{B'}{KM^+c_{max}}\rfloor$, r^+ with $\lfloor\frac{B'}{KM^-c_{max}}\rfloor$, and let $w^+=1$, we have

$$R_\alpha < \frac{B'M^+}{M^-c_{min}}+\frac{2M^+B}{M^-c_{min}}\cdot\sqrt{\frac{\delta NM^+c_{max}\ln(\frac{M^+B}{M^-c_{max}})}{B'M^-}} \quad (19)$$

When $B'=(\frac{1}{M^-})^{\frac{1}{3}}(\delta\cdot NM^+c_{max}\ln(\frac{M^+B}{M^-c_{max}}))^{\frac{1}{3}}B^{\frac{2}{3}}$, the above upper bound can achieve a minimum value:

$$\frac{3M^+}{M^-c_{min}}(\frac{1}{M^-})^{\frac{1}{3}}(\delta\cdot NM^+c_{max}\ln(\frac{M^+B}{M^-c_{max}}))^{\frac{1}{3}}B^{\frac{2}{3}} \quad (20)$$

Therefore, the theorem holds. □

5.3 Regret of ACMABA

Based on the above analysis, we denote the winner sets produced by the α-optimal and ACMABA algorithms in each round as $\boldsymbol{V}'=\{v_1,\cdots,v_K\}$ and $\boldsymbol{S}'_t=\{s_1,\cdots,s_K\}$, respectively. Then, we define the smallest and largest possible differences of RCR value and the largest possible difference of revenue among all non-optimal winner sets $\boldsymbol{S}'_t\neq\boldsymbol{V}'$:

$$\Delta_{max}=\sum_{i\in\boldsymbol{V}'}\frac{\sum_{j\in\mathcal{M}_i}w_j q_i}{b_i}-\min_{\boldsymbol{S}'_t\neq\boldsymbol{V}'}\sum_{i\in\boldsymbol{S}'_t}\frac{\sum_{j\in\mathcal{M}_i}w_j q_i}{b_i}, \quad (21)$$

$$\Delta_{min}=\sum_{i\in\boldsymbol{V}'}\frac{\sum_{j\in\mathcal{M}_i}w_j q_i}{b_i}-\max_{\boldsymbol{S}'_t\neq\boldsymbol{V}'}\sum_{i\in\boldsymbol{S}'_t}\frac{\sum_{j\in\mathcal{M}_i}w_j q_i}{b_i}, \quad (22)$$

$$\varrho^q_{max}=\sum_{i\in\boldsymbol{V}'}\sum_{j\in\mathcal{M}_i}w_j q_i-\min_{\boldsymbol{S}'_t\neq\boldsymbol{V}'}\sum_{i\in\boldsymbol{S}'_t}\sum_{j\in\mathcal{M}_i}w_j q_i. \quad (23)$$

Let $\gamma_{i,t}$ be the counter of worker i after the initial exploration. In each round t after the initial exploration, when $\boldsymbol{S}'_t\neq\boldsymbol{V}'$, the counter $\gamma_{i,t}$ is updated as follows:

$$i=\arg\min_{i'\in\boldsymbol{S}'_t}\gamma_{i',t-1}, \qquad \gamma_{i,t}=\gamma_{i,t-1}+1. \quad (24)$$

That is, we find the smallest counter of all recruited workers at the end of $(t-1)$-th round. If multiple workers satisfy the condition, we randomly select any one. So the worker with the smallest counter $\gamma_{i,t}$ will be incremented by 1. This means that for $\forall i\in\mathcal{N}$, the sum of the counter $\gamma_{i,t}$ equals to the total times that the sub-optimal winner set to be recruited. When any non-optimal workers set is determined in a round, there is exactly one worker's counter to be incremented. Moreover, we let $\mu_{i,t}$ denote the exact times that worker i has been recruited until t-th round, where $\gamma_{i,t}\leq\mu_{i,t}$. Next, we will focus on the upper bound of the counter $\gamma_{i,T}$, where T is the total rounds of the ACMABA algorithm under the budget constraint B.

Lemma 16. *The expected counter γ_i^T has an upper bound for any worker $i\in\mathcal{N}$ under the budget constraint B, that is*

$$E[\gamma_i^T]\leq\eta_1\ln(T)+\eta_2, \quad (25)$$

where $\eta_1=\begin{cases}\frac{4K^2\delta}{(c_{min}\Delta_{min})^2}+2K, \delta=K+1/2\\ \frac{4K^2\delta}{(c_{min}\Delta_{min})^2}, \delta>K+1/2\end{cases}$,

$\eta_2=\begin{cases}\frac{4K^2\delta\ln(KM^+)}{(c_{min}\Delta_{min})^2}+1+2K, \delta=K+1/2\\ \frac{4K^2\delta\ln(KM^+)}{(c_{min}\Delta_{min})^2}+1+2K\cdot O(1), \delta>K+1/2\end{cases}$.

Proof. In each round t, one of the following cases must happen: 1) the optimal set of workers, *i.e.*, $\boldsymbol{V}'$, might be selected; 2) a non-optimal set of workers will be selected, *i.e.*, $\boldsymbol{S}'_t\neq\boldsymbol{V}'$. In the first case, the counter $\gamma_{i,t}$ will not change, while in the second case, the counter $\gamma_{i,t}$ will be updated according to Eq. (24). Recall that we use $\phi_{i,t}\in\{0,1\}$ to denote the selection policy for worker i in the t-th round. That is, $\phi_{i,t}$ represents the change of the counter $\gamma_{i,t}$, where $\phi_{i,t}=1$ means that $\gamma_{i,t}$ is incremented, and $\phi_{i,t}=0$ otherwise. Assume that $T'=\lceil\frac{N}{K}\rceil+1$ is the first round after initial exploration. Based on this, we have the following results:

$$\begin{aligned} \gamma_i^T &= \textstyle\sum_{t=T'}^T \boldsymbol{\phi}\{\phi_{i,t}=1\}\leq\ell+\sum_{t=1}^T\boldsymbol{\phi}\{\phi_{i,t}=1,\gamma_{i,t}\geq\ell\} \\ &\leq \ell+\textstyle\sum_{t=1}^T\boldsymbol{\phi}\Big\{\sum_{i\in\boldsymbol{S}'_t}RCR_{i,t+1}\geq\sum_{i\in\boldsymbol{V}'}RCR_{i,t+1},\gamma_{i,t}\geq\ell\Big\} \end{aligned}$$

$$\leq \ell + \sum_{t=1}^{T} \phi\Big\{ \max_{\ell\leq\mu_{s_1,t}\leq\cdots\leq\mu_{s_K,t}\leq t} RCR_{s_l,t} \geq \min_{1\leq\mu^*_{s_1}\leq\cdots\leq\mu^*_{s_K}\leq t} RCR_{v_l,t} \Big\}$$
$$\leq \ell + \sum_{t=1}^{T} \sum_{\mu_{s_1,t}=\ell}^{t} \cdots \sum_{\mu_{s_K,t}=\ell}^{t} \sum_{\mu_{v_1,t}=1}^{t} \cdots \sum_{\mu_{v_K,t}=1}^{t} \phi\Big\{ \sum_{l=1}^{K} RCR_{s_l,t} \geq \sum_{l=1}^{K} RCR_{v_l,t} \Big\}, \quad (26)$$

Then, we focus on the bound of $\sum_{l=1}^{K} RCR_{s_l,t} \geq \sum_{l=1}^{K} RCR_{v_l,t}$. According to Eq. (11):

$$\sum_{l=1}^{K} \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j(\hat{q}_{s_l,t}+\varepsilon_{s_l,t})}{b_{s_l}} \geq \sum_{l=1}^{K} \frac{\sum_{j\in\mathcal{M}_{v_l}} w_j(\hat{q}_{v_l,t}+\varepsilon_{v_l,t})}{b_{v_l}}. \quad (27)$$

We can obtain that at least one of the following cases must be true (which is based on the proof by contradiction):

$$\sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{v_l}} w_j\hat{q}_{v_l,t}/b_{v_l} \leq \sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{v_l}} w_j(q_{v_l}-\varepsilon^t_{v_l})/b_{v_l}, \quad (28)$$
$$\sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{s_l}} w_j\hat{q}_{s_l,t}/b_{s_l} \geq \sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{s_l}} w_j(q_{s_l}+\varepsilon^t_{s_l})/b_{s_l}, \quad (29)$$
$$\sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{v_l}} w_j q_{v_l}/b_{v_l} < \sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{s_l}} w_j(q_{s_l}+2\varepsilon^t_{s_l})b_{s_l}. \quad (30)$$

Next, we need to prove the upper bound of the probability of Eqs. (28) and (29). By applying the Chernoff-Hoeffding bound in Lemma 12, we can get

$$\mathbb{P}\Big\{\sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{v_l}} \frac{w_j\hat{q}_{v_l,t}}{b_{v_l}} \leq \sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{v_l}} \frac{w_j(q_{v_l}-\varepsilon^t_{v_l})}{b_{v_l}}\Big\}$$
$$\leq \sum_{l=1}^{K} \mathbb{P}\{\hat{q}_{v_l,t} \leq (q_{v_l}-\varepsilon^t_{v_l})\} \leq \sum_{l=1}^{K} e^{-2n^t_{v_l}\varepsilon^{t\,2}_{v_l}}$$
$$= \sum_{l=1}^{K} (\sum_{i\in\mathcal{N}} n_{i,t})^{-2\delta} \leq K(tKM^-)^{-2\delta} \leq Kt^{-2\delta}. \quad (31)$$

We can also similarly prove that

$$\mathbb{P}\Big\{\sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{s_l}} \frac{w_j\hat{q}_{s_l,t}}{b_{s_l}} \geq \sum_{l=1}^{K}\sum_{j\in\mathcal{M}_{s_l}} \frac{w_j(q_{s_l}+\varepsilon^t_{s_l})}{b_{s_l}}\Big\}$$
$$\leq K(tKM^-)^{-2\delta} \leq Kt^{-2\delta}. \quad (32)$$

Then, we choose a certain value ℓ to make the Eq. (30) impossible. Based on the fact that $\mu^t_i \geq \gamma_{i,t} \geq \ell$, we have

$$\sum_{l=1}^{K} \frac{\sum_{j\in\mathcal{M}_{v_l}} w_j q_{v_l}}{b_{v_l}} - \sum_{l=1}^{K} \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j q_{s_l}}{b_{s_l}} - 2\sum_{l=1}^{K} \frac{\sum_{j\in\mathcal{M}_{s_l}} w_j \varepsilon_{s_l,t}}{b_{s_l}}$$
$$\geq \Delta_{min} - 2\sum_{l=1}^{K} |M_{s_l}| w^+ \sqrt{\frac{\delta\ln(\sum_{i\in\mathcal{N}} n_{i,t})}{n_{s_l,t}}} / b_{s_l}$$
$$\geq \Delta_{min} - 2w^+ \sum_{l=1}^{K} \sqrt{\frac{\delta\ln(\sum_{i\in\mathcal{N}} n_{i,t})}{|\mathcal{M}_{s_l}|\ell}} / c_{min} \geq 0. \quad (33)$$

After analyzing Eq. (33), we can yield that Eq. (33) always holds if ℓ satisfies the following condition:

$$\ell \geq \frac{4K^2\delta\ln(TKM^+)}{(c_{min}\Delta_{min})^2}. \quad (34)$$

Now we continue Eq. (26), and get

$$E[\gamma^T_i] \leq \Big\lceil \frac{4K^2\delta\ln(TKM^+)}{(c_{min}\Delta_{min})^2} \Big\rceil + \sum_{t=1}^{T} (t-\ell+1)^K t^K 2Kt^{-2\delta}$$
$$\leq \frac{4K^2\delta\ln(TKM^+)}{(c_{min}\Delta_{min})^2} + 1 + 2K\sum_{t=1}^{T} t^{-2(\delta-K)} \quad (35)$$

In ACMABA, we set $\delta \geq (K+1/2)$. Then we can get

$$E[\gamma^T_i] \leq \frac{4K^2\delta\ln(TKM^+)}{(c_{min}\Delta_{min})^2} + 1 + \begin{cases} 2K(ln(T)+1), & \delta=K+1/2 \\ 2K\cdot O(1), & \delta>K+1/2 \end{cases}$$
$$\leq \frac{4K^2\delta\big(\ln(T)+\ln(KM^+)\big)}{(c_{min}\Delta_{min})^2} + 1 + \begin{cases} 2K(ln(T)+1), & \delta=K+1/2 \\ 2K\cdot O(1), & \delta>K+1/2 \end{cases}$$
$$= \eta_1\ln(T) + \eta_2, \quad (36)$$

where $\eta_1 = \begin{cases} \frac{4K^2\delta}{(c_{min}\Delta_{min})^2} + 2K, & \delta=K+1/2 \\ \frac{4K^2\delta}{(c_{min}\Delta_{min})^2}, & \delta>K+1/2 \end{cases}$,
$\eta_2 = \begin{cases} \frac{4K^2\delta\ln(KM^+)}{(c_{min}\Delta_{min})^2} + 1 + 2K, & \delta=K+1/2 \\ \frac{4K^2\delta\ln(KM^+)}{(c_{min}\Delta_{min})^2} + 1 + 2K\cdot O(1), & \delta>K+1/2 \end{cases}$.
Hence, the lemma holds. □

Since the bound of $\gamma_{i,T}$ is determined by expected total rounds T in the ACMABA algorithm, we then analyze the bounds of T under the limited budget B. Based on the above analysis, the payment for each winning worker i in the α-optimal algorithm is $p_i = \frac{RCR_i}{RCR_{K+1}} b_{K+1} \geq b_i$. We denote the total payment in each round as follows:

$$P_{\boldsymbol{V}'} = \sum_{i\in\boldsymbol{V}'} \frac{RCR_i}{RCR_{K+1}} b_{K+1} \geq \sum_{i\in\boldsymbol{V}'} b_i \quad (37)$$

We can derive the bounds of expected total rounds T.

Lemma 17. *The the expected total rounds T of the ACMABA algorithm is bounded as follows:*

$$\frac{B}{P_{\boldsymbol{V}'}} - \eta_1\eta_4\ln(\frac{2B}{P_{\boldsymbol{V}'}} + \eta_3) - \eta_2\eta_4 - 1 \leq T \leq \frac{2B}{P_{\boldsymbol{V}'}} + \eta_3, \quad (38)$$

where $\eta_3 = \frac{2NM^+c_{max}}{KM^-c_{min}}\big(\eta_1\ln(\frac{2NM^+c_{max}\eta_1}{KM^-c_{min}}) - \eta_1 + \eta_2\big)$ *and* $\eta_4 = \frac{NM^+c_{max}}{P_{\boldsymbol{V}'}}$.

Proof. Here, we let $T(B)$ denote the total rounds under the budget B in the proof, which equals to T. Let $T'(B)$ denote the total rounds of the α-optimal algorithm under the budget B. Since the workers' qualities are known in the α-optimal algorithm and the bids submitted by workers are fixed, the optimal set of winning workers in each round is determined, i.e., $\boldsymbol{V}'$. Then, the payment for each winning worker can be determined and the total payment in one round can be calculated according to Eq. (37). Thus, the value of total rounds $T'(B)$ is fixed, i.e., $T'(B) = \lfloor B/P_{\boldsymbol{V}'} \rfloor$.

To derive the upper bound of $T(B)$, we first analyze the relationship between $T(B)$ and $T'(B)$. According to Eq. (13), the payment for each winner in ACMABA is greater than its true cost. Thus, the total payment in each round is greater than $K\cdot c_{min}$, and we have $T \leq B/Kc_{min}$. Then,

$$T(B) \leq T'(B) + T(\sum_{i\notin\boldsymbol{V}'} \mu_{i,T(B)} \cdot M^+ \cdot c_{max})$$
$$\leq T'(B) + M^+c_{max}T(\sum_{i=1}^{N} \gamma_{i,T(B)})$$
$$\leq \frac{B}{P_{\boldsymbol{V}'}} + \frac{NM^+c_{max}}{KM^-c_{min}}\big(\eta_1\ln(T(B)) + \eta_2\big). \quad (39)$$

According to the inequality $\ln x \leq x-1$, for $\forall x>0$, we have

$$\ln(\frac{KM^-c_{min}}{2NM^+c_{max}\eta_1}) \leq \frac{KM^-c_{min}}{2NM^+c_{max}\eta_1} - 1 \quad (40)$$
$$\Rightarrow \ln(T(B)) \leq \frac{KM^-c_{min}}{2NM^+c_{max}\eta_1} - 1 + ln(\frac{2NM^+c_{max}\eta_1}{KM^-c_{min}}). \quad (41)$$

By substituting the bound of $\ln(T(B))$ in Eq. (39) with Eq. (41), we can have

$$T(B) \leq \frac{B}{P_{\boldsymbol{V}'}} + \frac{T(B)}{2} + \frac{NM^+c_{max}}{KM^-c_{min}}\big(\eta_1\ln(\frac{2NM^+c_{max}\eta_1}{KM^-c_{min}}) - \eta_1 + \eta_2\big)$$
$$\Rightarrow T(B) \leq \frac{2B}{P_{\boldsymbol{V}'}} + \frac{2NM^+c_{max}}{KM^-c_{min}}\big(\eta_1\ln(\frac{2NM^+c_{max}\eta_1}{KM^-c_{min}}) - \eta_1 + \eta_2\big). \quad (42)$$

Let $\eta_3 = \frac{2NM^+c_{max}}{KM^-c_{min}}\big(\eta_1\ln(\frac{2NM^+c_{max}\eta_1}{KM^-c_{min}}) - \eta_1 + \eta_2\big)$ for simplicity. Then, we have $T(B) \leq \frac{2B}{P_{\boldsymbol{V}'}} + \eta_3$.

Next, we focus on the lower bound of $T(B)$. Here, we divide the budget B into two parts: B' and B^-, in which B' is used to recruit the optimal set of workers and B^-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing

indicates the remaining budget spent in recruiting the non-optimal sets of workers. Hence, we can get $T(B')\leq T'(B)$ and $B/P_{\boldsymbol{V}'}-1\leq T'(B)\leq B/P_{\boldsymbol{V}'}$. Since $T'(B)$ and $T(B)$ are based on the same payment computation method, we have

$$\begin{aligned}
&T(B)=T(B'+B^-)\geq T'(B')\\
&\geq T'(B-\textstyle\sum_{i\notin \boldsymbol{V}'}\mu_{i,T(B)}\cdot M^+\cdot c_{max})\\
&\geq T'(B-M^+c_{max}\textstyle\sum_{i=1}^N\gamma_{i,T(B)})\\
&\geq T'(B-NM^+c_{max}(\eta_1\ln(T(B)))+\eta_2)\\
&\geq(B-NM^+c_{max}(\eta_1\ln(T(B)))+\eta_2)/P_{\boldsymbol{V}'}-1. \qquad (43)
\end{aligned}$$

Then, we substitute the logarithm of Eq. (42) into Eq. (43) to derive the lower bound of $T(B)$:

$$\begin{aligned}
T(B)\geq&\tfrac{B}{P_{\boldsymbol{V}'}}-\tfrac{NM^+c_{max}(\eta_1\ln(2B/P_{\boldsymbol{V}'}+\eta_3)+\eta_2)}{P_{\boldsymbol{V}'}}-1\\
\geq&\tfrac{B}{P_{\boldsymbol{V}'}}-\eta_1\eta_4\ln(\tfrac{2B}{P_{\boldsymbol{V}'}}+\eta_3)-\eta_2\eta_4-1, \qquad (44)
\end{aligned}$$

where $\eta_4=\frac{NM^+c_{max}}{P_{\boldsymbol{V}'}}$. Therefore, the lemma holds. □

Finally, we prove the upper bound of the expected regret for ACMABA, and we have the following theorem.

Theorem 18. *The expected regret of the ACMABA algorithm is bounded by* $O\big(NMK^3\ln(B+NMK^2\ln(NMK^2))\big)$.

Proof. Let the total revenue obtained by the α-optimal algorithm in each round be $R_{\boldsymbol{V}'}$. According to the definition of regret in Eq. (14) and Lemmas 16 and 17, we have:

$$\begin{aligned}
R_\alpha=&\tfrac{B}{P_{\boldsymbol{V}'}}R_{\boldsymbol{V}'}-T(B)R_{\boldsymbol{V}'}+T(B)R_{\boldsymbol{V}'}-E[\sum_{t=1}^{T}\sum_{i\in \boldsymbol{S}'_t}\sum_{j\in\mathcal{M}_i}w_jq_i]\\
\leq&\ \tfrac{B}{P_{\boldsymbol{V}'}}R_{\boldsymbol{V}'}-T(B)R_{\boldsymbol{V}'}+\textstyle\sum_{i=1}^N\gamma_{i,T}\varrho^q_{max}\\
\leq&\ \tfrac{B}{P_{\boldsymbol{V}'}}R_{\boldsymbol{V}'}-(\tfrac{B}{P_{\boldsymbol{V}'}}-\eta_1\eta_4\ln(\tfrac{2B}{P_{\boldsymbol{V}'}}+\eta_3)-\eta_2\eta_4-1)R_{\boldsymbol{V}'}\\
&+N\varrho^q_{max}(\eta_1\ln(\tfrac{2B}{P_{\boldsymbol{V}'}}+\eta_3)+\eta_2)\\
\leq&\ (N\varrho^q_{max}+\eta_4R_{\boldsymbol{V}'})\eta_1\ln(\tfrac{2B}{P_{\boldsymbol{V}'}}+\eta_3)+\eta_2\eta_4R_{\boldsymbol{V}'}\\
&+\eta_2N\varrho^q_{max}+R_{\boldsymbol{V}'}\\
=&\ O\big(NM^+K^3\ln(B+\tfrac{NM^+K^2}{M^-}\ln(\tfrac{NM^+K^2}{M^-}))\big). \qquad (45)
\end{aligned}$$

Therefore, the theorem holds. □

5.4 Truthfulness, Individual Rationality and Efficiency

To prove the truthfulness, we reveal that workers can obtain the maximum utility if they bid truthfully.

Theorem 19. *The CMABA mechanism is truthful.*

Proof. For $\forall$ worker $i\in\mathcal{N}$ submits an untruthful bid b_i to complete tasks in $\mathcal{M}_i$, i.e., $b_i\neq c_i$. In the exploration phase, the task allocation is irrespective of the bid and the payment is fixed ($|\mathcal{M}_i|\,c_{max}$) when i is recruited. Thus, the property is satisfied because the worker cannot increase its utility by manipulating the real cost. In the exploitation phase, in each round t, we denote the recruitment result for bidding b_i and c_i as $\phi'_{i,t}$ and $\phi_{i,t}$, respectively. There are two cases:

Case 1 ($b_i>c_i$): According to Algorithm 1, we can derive that $\phi'_{i,t}\leq\phi_{i,t}$. If $\phi'_{i,t}=\phi_{i,t}$, the property is trivially satisfied as the payment does not depend on the bid of i; if $\phi'_{i,t}<\phi_{i,t}$, then $\phi'_{i,t}=0$ and $\phi_{i,t}=1$, bidding truthfully is a better choice.

Case 2 ($b_i<c_i$): We have $\phi'_{i,t}\geq\phi_{i,t}$. If $\phi'_{i,t}=\phi_{i,t}$, the property is satisfied; if $\phi'_{i,t}>\phi_{i,t}$, then $\phi'_{i,t}=1$ and $\phi_{i,t}=0$. This implies that $b_i\leq\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}}$ and $c_i\geq\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}}$. So we can derive $\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}}\leq c_i\leq|\mathcal{M}_i|c_{max}$. The utility of worker i is $u_{i,t}(b_i)=p_{i,t}(b_i)-c_i=\min\{\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}},|\mathcal{M}_i|c_{max}\}-c_i\leq 0$.

Therefore, it is a dominant strategy for each worker to report its true sensing cost, which proves the theorem. □

Note that the bids remain unchanged in any round of CMABA and ACMABA. In CMABA, the reverse auction will be executed once at the beginning of exploitation, using the derived UCB values at the end of exploration. Since the UCB indexes are fixed values for CMABA, truthfulness holds in the exploitation phase according to Theorem 19. Similarly, in ACMABA, truthfulness holds in the second round (i.e., the first round after initial exploration). However, due to the variance of each worker i's quality, worker i's sensing quality varies continuously in each following round. The varied qualities will bring about estimation error of the sample mean value $\hat{q}_i^t$, and even cause deviation on the worker recruitment result and the critical payment, which makes ACMABA achieve the approximate truthfulness.

Theorem 20. *In each round t, the CMABA and ACMABA mechanisms are individually rational.*

Proof. In the t-th round, if a worker $i\in\mathcal{N}$ is not recruited, its utility is zero. Otherwise, if i is recruited , its utility is $u_{i,t}=p_{i,t}-c_i$. In the exploration phase, the payment of i is always $|\mathcal{M}_i|\,c_{max}\geq c_i$, so worker i's utility is nonnegative. In the exploitation phase, if worker i is recruited, we can derive that $b_i\leq\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}}$. Based on Theorem 19, each worker will bid truthfully, i.e., $b_i=c_i$. Hence, $c_i\leq\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}}$. Then, the utility of worker i is $\min\{\frac{\sum_{j\in\mathcal{M}_i}w_j\hat{q}_i^+}{RCR_{s_{K+1}}},|\mathcal{M}_i|c_{max}\}-c_i\geq 0$. □

Theorem 21. *The CMABA and ACMABA mechanisms are computationally efficient.*

Proof. The CMABA and ACMABA mechanisms consist of the exploration and exploitation phases. In each iteration of the while loop in the exploration phase, the computational complexity is at most $O(N)$ dominated by the quality update. In the exploitation phase, there are two main operations of sorting and payment, whose total computational complexity is $O(N\log N)$. The iteration of recording bandit policy is an order of $O(K)$. In CMABA, the while loops of exploration and exploitation phases are executed at most $\frac{B}{KM^-c_{max}}$ and $\frac{B}{KM^-c_{max}}$ times, respectively. In ACMABA, the while loops equals to the total rounds T which is bounded in Theorem 17. Hence, CMABA and ACMABA are computationally efficient. □

6 Performance Evaluation

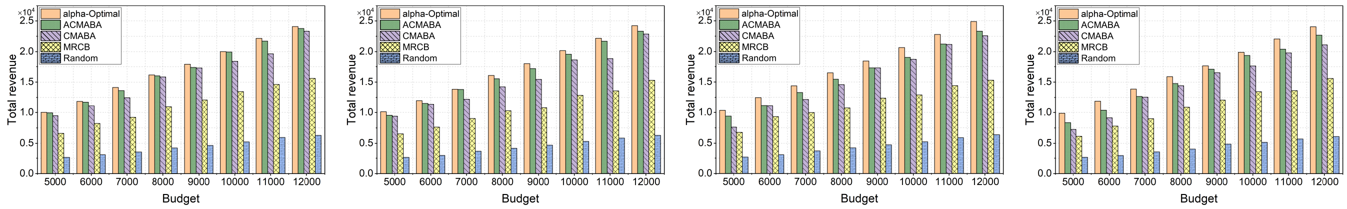
In this section, we evaluate the performance of CMABA and ACMABA through simulations on real-world data traces.

6.1 Evaluation Methodology

6.1.1 Simulation Setup

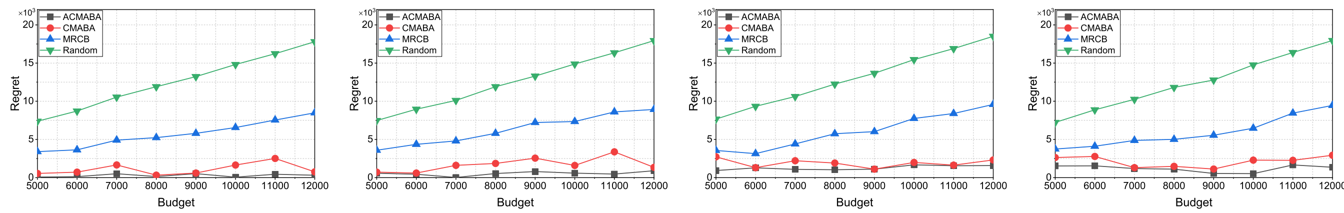
We conduct extensive simulations based on Chicago Taxi Trips [26] including 27465 taxi records. Each entry of the trace records the taxiID, time stamp, trip miles and the location of picking up/dropping off passengers, etc. In our

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing



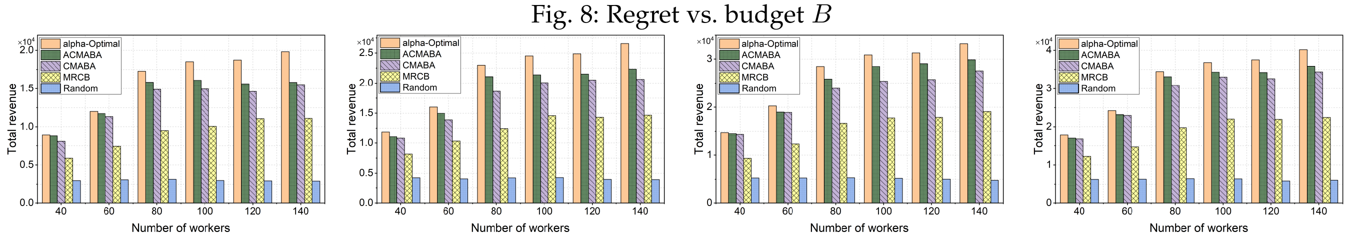
(a) $N=40$ (b) $N=60$ (c) $N=80$ (d) $N=100$

Fig. 7: Total revenue vs. budget B



(a) $N=40$ (b) $N=60$ (c) $N=80$ (d) $N=100$

Fig. 8: Regret vs. budget B



(a) $B=6000$ (b) $B=8000$ (c) $B=10000$ (d) $B=12000$

Fig. 9: Total revenue vs. the number of workers N

TABLE 2: Simulation settings

Parameter name	Values
Budget B	5, 6, 7, 8, 9, **10**, 11, 12 ($\times 10^3$)
number of tasks M, $\lvert\mathcal{M}_i\rvert$	200, [5, 15]
number of worker N	40, 60, 80, 100 120, **140**
percentage of workers K/N	16%, 20%, 25%, **33%**, 50%, 60%
range of cost $[c_{min}, c_{max}]$	[0.1, 1]
range of expected quality q_i	[0, 1]

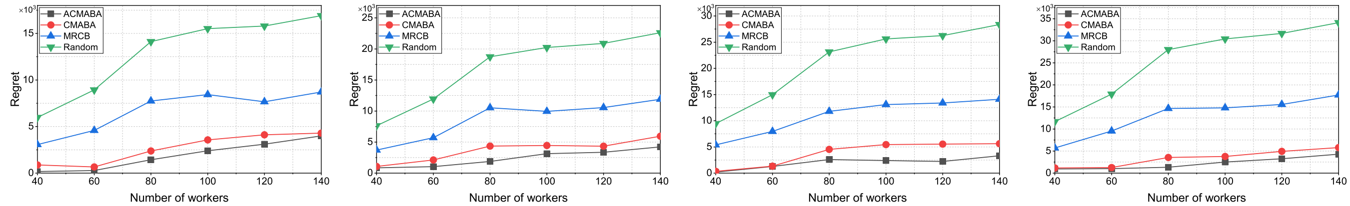
simulation, we regard taxis as sensing workers and assume that a taxi can perform the sensing tasks at the locations where it picks up or drops off passengers. We first choose M locations and N taxis from the trace, where $M = 200$ and N is selected from $[40, 140]$. The number of tasks $|\mathcal{M}_i|$ that a worker can perform is randomly selected from $[5, 15]$, which means that the number of candidate locations that the corresponding taxi can cover ranges from 5 to 15. In addition, the weight of tasks is uniform, and we set $K = \lfloor N/3 \rfloor$ as default. We assume that the cost for each task is proportional to the trip miles and let the cost for each task not exceed $[0.1, 1]$. Since there is no record about the qualities, we adopt truncated Gaussian distribution to generate workers' expected qualities from $[0, 1]$.

6.1.2 Algorithms for Comparison

Since our two mechanisms consider both unknown qualities and incentive issues in MCS, there is no existing algorithm that can be directly applied to our problem. To the best of our knowledge, the closest algorithm that can be adapted to our setting is the MRCB algorithm [20], which is a budget-limited K-armed bandit algorithm. However, since MRCB assumes that both the rewards and costs are stochastic but neglects the incentive issues, we need to incorporate reverse auction into MRCB to deal with the strategic workers in our model. More specifically, we split the total budget into two equal parts to achieve truthfulness, which are used for exploration and exploitation, respectively. Moreover, we implement the α-Optimal algorithm and the Random algorithm for better comparison. The α-Optimal algorithm has full knowledge about workers' qualities and always recruits the best K workers. The Random algorithm randomly selects K workers in each round and pays the winning workers the maximum cost they might incur, i.e., $|\mathcal{M}_i|c_{max}$.

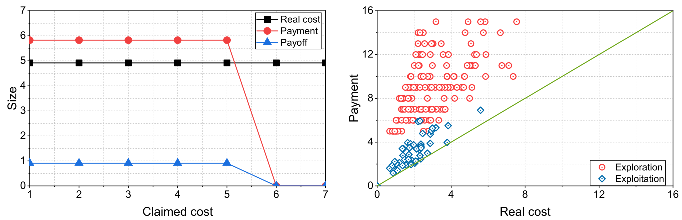
6.2 Evaluation Results

First, we investigate the impact of budget B from 5000 to 12000 with an increment of 1000 and exhibit the performance of different algorithms in terms of total revenue and regret, as shown in Figs. 7 and 8, respectively. In Fig. 7, we can observe that ACMABA and CMABA both achieve much higher total revenue than the MRCB and Random algorithms. More specifically, the total revenue achieved by CMABA is at least 45% higher than MRCB and is almost three times of Random algorithm, while ACMABA achieves a little higher total revenue than CMABA. Actually, the total revenue of CMABA and ACMABA is even going to catch up with the α-Optimal algorithm which knows the quality information in advance. In Fig. 8, the results show that the regrets of ACMABA and CMABA are much lower than the MRCB and Random algorithms. Since more sensing workers can be recruited under a larger budget, the total revenue and regret of all algorithms generally increase with the increasing budget. The small fluctuations of regret in ACMABA and CMABA because the observed quality of each worker in different rounds may be a little higher or lower than the expected quality. Basically, the regrets of ACMABA and CMABA are small due to the quality learning.



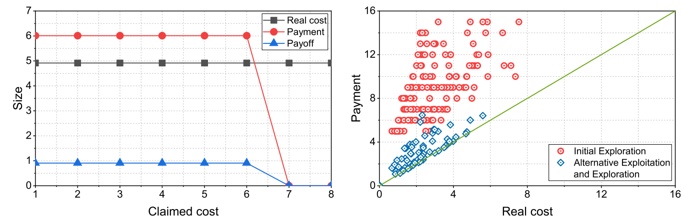
(a) $B=6000$ (b) $B=8000$ (c) $B=10000$ (d) $B=12000$

Fig. 10: Regret vs. the number of workers N



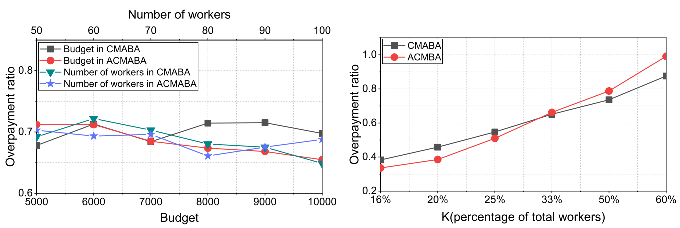
(a) Truthfulness (b) Individual Rationality

Fig. 11: Auction of CMABA



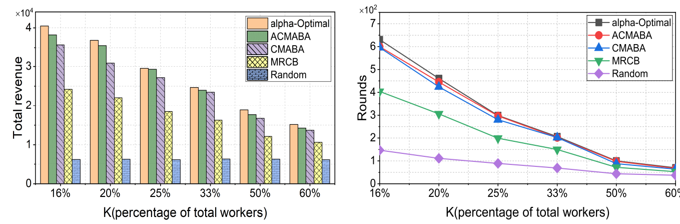
(a) Truthfulness (b) Individual Rationality

Fig. 12: Auction of ACMABA



(a) λ vs. Budget and N (b) λ vs. K

Fig. 13: Overpayment ratio in CMABA and ACMABA



(a) Total revenue vs. K (b) Total rounds vs. K

Fig. 14: Evaluation on the parameter K

Second, we evaluate the performances of ACMABA and CMABA with the other algorithms by changing the number of workers, as depicted in Figs. 9-10, where $K=20$. The figures show that ACMABA and CMABA can achieve much higher total revenue and much lower regret than the MRCB and Random algorithms, where ACMABA performs better than CMABA. The total revenue of all mechanisms exhibits an uptrend with the increasing N. This is because when we fix the budget B and K, and let more candidate workers emerge, there are more workers with higher qualities and lower costs, leading to a better selection than before. When N is much larger than K, more budget is required either to explore the workers in CMABA or to take more chances to recruit workers in ACMABA. Hence, the regrets of CMABA and ACMABA both increase with the increasing N.

To verify the properties of the CMABA and ACMABA mechanisms, we consider the three metrices: truthfulness, individual rationality and overpayment ratio.

Truthfulness. We randomly pick a winning bid, change its claimed cost, and recalculate the related payment as well as the utility. The results, as depicted in Figs. 11(a) and 12(a) to present CMABA and ACMABA respectively, both show that the utility remains unchanged if the claimed cost is no more than the payment. However, when the claimed cost is larger than the critical value, the utility becomes zero. Thus, the CMABA and ACMABA mechanisms satisfy truthfulness.

Individual rationality. We set $N=140$ and $K=\lfloor \frac{N}{3} \rfloor$, then calculate the payments for workers in the exploration and exploitation phases, respectively. Figs. 11(b) and 12(b) show that each payment is no less than the corresponding real cost, which demonstrates that CMABA and ACMABA achieve individual rationality. Additionally, we can see that workers are paid c_{max} in the (initial) exploration phase in CMABA and ACMABA, most of which are much higher than the costs (i.e., the red symbols). But the payment for each worker is critical and relatively close to the cost in the exploitation phase of CMABA, as well as the alternative exploitation and exploration phase in ACMABA.

Overpayment ratio. The evaluation results of overpayment ratio are depicted in Fig. 13, which can be computed by

$$\lambda = \frac{\sum_t \sum_{i \in \mathcal{N}} (p_{i,t} - c_i)\phi_{i,t}}{\sum_t \sum_{i \in \mathcal{N}} c_i \phi_{i,t}}. \quad (46)$$

In Fig. 13(a), we show the overpayment ratio of CMABA and ACMABA when varying the budget from 5000 to 10000 and varying the number of workers from 50 to 100. We can find that λ is stable and is always less than 0.75, which means that the requester does not have to pay much extra money to induce truthfulness with the fix ratio of $K/N \approx 33\%$. In Fig. 13(b), the overpayment ratio increases significantly when the ratio of K/N increases whatever in CMABA or in ACMABA. So the selection of K value is important in our two mechanisms, which inspires us to explore the impact of the parameter K on our mechanisms.

In Fig. 14, we present the changes of total revenue and execution rounds when K varies with the given B=14000 and $N=140$. The results indicate that ACMABA and CMABA still achieve much higher revenue than the MRCB and Random algorithms within different K. The smaller K means that the higher total sensing revenue can be achieved. This can be explained by the reason that a smaller K allows less suboptimal workers to be recruited in each round. However, this will also result in more recruitment rounds.

7 Related Work

We review the related work from the following aspects:

Incentive Mechanism: Incentive mechanisms have been widely investigated in many fields, especially for the worker recruitment problem in MCS. There are various works on online workers [7], [8], data quality management [13], [14], privacy preservation [27], data publishing [28] and so on. Different optimization objectives have also been considered, such as quality maximization [29], cost minimization [30]. However, most of these works assume that workers' quality information is known in advance.

Learning-based Approach: [2] investigates many researches that further optimize MCS by integrating learning techniques to infer the unknown knowledge, such as workers' qualities, workers behavioral patterns, sensing data correlation, etc. For example, [31], [32] consider the possibility of worker's rejection for tasks, which is predicted by two methods, maximum likelihood estimation and weighted utility sum in [31] and multiclassifier based approach in [32], respectively. [33] extracts topology features by supervised training and uses the Back-Propagation neural network learning algorithm to minimize computation overhead. And [34] uses multi-population co-evolution to enhance the robustness of scale-free topologies. [35], [36] take account of the correlation among data to improve performance of learning model. [37] calculates the probability of participants' presence in each spatial-temporal cell based on historical trajectories. [38] reviews many quality-aware projects and establishes a categorization based on the characteristics of data, tasks, and scopes of projects to design quality learning strategies. However, all of the above learning-based approaches are based on the historical records or some priori knowledge, while in our MCS scenario, the system completely does not have any prior information of the workers' qualities.

MAB Mechanism: Multi-armed bandit is an efficient reinforcement learning model to make online decisions under information unknown environments. Some recent works in MCS [9], [9], [23] adopt the multi-armed bandit model to consider the scenario where workers' qualities or costs are unknown a priori. Nevertheless, these works either consider only one task or do not take the incentive issues into account. There are also a few works with incentive mechanism. For instance, [39] designs an MAB mechanism to incentivize electricity reduction in smart grids for T rounds. [40] proposes a budget-limited MAB mechanism and proves that the regret bound for any truthful budgeted MAB mechanism is $\Omega(N^{\frac{1}{3}}B^{\frac{2}{3}})$. [41] develops a mechanism that achieves approximate truthfulness and individual rationality. However, these researches only consider the problem of pulling a single arm in each round and assume the task sets for all workers are identical, which is not always supported in MCS. [42] designs an incentive mechanism to recruit a group of workers in each round based on Thompson Sampling, without considering the budget constraint.

8 Conclusion

In this paper, we focus on the worker recruitment problem in MCS while considering the unknown qualities and strategic workers simultaneously. We model the worker recruitment problem as a novel K-armed combinatorial multi-armed bandit problem. Moreover, we adopt reverse auction to incentivize workers' participation as well as discourage their strategic behaviors. Then, we propose two incentive mechanisms called CMABA and ACMABA, which tackle the trade-off of exploration and exploitation with theoretical analysis, and achieve truthfulness, individual rationality and computational efficiency. Finally, we analyze the worst regret bounds and conduct extensive simulations on a real trace to demonstrate the effectiveness of our mechanisms.

References

[1] M. Xiao, J. Wang, H. Zhao, and G. Gao, "Unknown worker recruitment in mobile crowdsensing using cmab and auction," in *IEEE ICDCS*, 2020.

[2] J. Wang, Y. Wang, D. Zhang, J. Goncalves, D. Ferreira, A. Visuri, and S. Ma, "Learning-assisted optimization in mobile crowd sensing: A survey," *IEEE Trans. Ind. Informatics*, vol. 15, no. 1, pp. 15–22, 2019.

[3] J. Wang, F. Wang, Y. Wang, D. Zhang, L. Wang, and Z. Qiu, "Social-network-assisted worker recruitment in mobile crowd sensing," *IEEE. Trans. Mob. Comput.*, vol. 18, no. 7, pp. 1661–1673, 2019.

[4] H. Jin, L. Su, and K. Nahrstedt, "Centurion: Incentivizing multi-requester mobile crowd sensing," in *IEEE INFOCOM*, 2017.

[5] G. Ding, J. Wang, Q. Wu, L. Zhang, Y. Zou, Y. Yao, and Y. Chen, "Robust spectrum sensing with crowd sensors," *IEEE Trans. Commun.*, vol. 62, no. 9, pp. 3129–3143, 2014.

[6] J. Wang, F. Wang, Y. Wang, D. Zhang, B. Lim, and L. Wang, "Allocating heterogeneous tasks in participatory sensing with diverse participant-side factors," *IEEE. Trans. Mob. Comput.*, vol. 18, no. 9, pp. 1979–1991, 2019.

[7] D. Zhao, X.-Y. Li, and H. Ma, "How to crowdsource tasks truthfully without sacrificing utility: online incentive mechanisms with budget constraint," in *IEEE INFOCOM*, 2014.

[8] M. Xiao, J. Wu, L. Huang, R. Cheng, and Y. Wang, "Online task assignment for crowdsensing in predictable mobile social networks," *IEEE. Trans. Mob. Comput.*, vol. 16, no. 8, pp. 2306–2320, 2017.

[9] G. Gao, J. Wu, M. Xiao, and G. Chen, "Combinatorial multi-armed bandit based unknown worker recruitment in heterogeneous crowdsensing," in *IEEE INFOCOM*, 2020.

[10] E. Wang, Y. Yang, J. Wu, W. Liu, and X. Wang, "An efficient prediction-based user recruitment for mobile crowdsensing," *IEEE. Trans. Mob. Comput.*, vol. 17, no. 1, pp. 16–28, 2018.

[11] Y. Yang, W. Liu, E. Wang, and J. Wu, "A prediction-based user selection framework for heterogeneous mobile crowdsensing," *IEEE. Trans. Mob. Comput.*, vol. 18, no. 11, pp. 2460–2473, 2019.

[12] X. Zhang and X. Gong, "Online data quality learning for quality-aware crowdsensing," in *IEEE SECON*, 2019.

[13] G. Gao, M. Xiao, J. Wu, L. Huang, and C. Hu, "Truthful incentive mechanism for nondeterministic crowdsensing with vehicles," *IEEE. Trans. Mob. Comput.*, vol. 17, no. 12, pp. 2982–2997, 2018.

[14] D. Peng, F. Wu, and G. Chen, "Data quality guided incentive mechanism design for crowdsensing," *IEEE. Trans. Mob. Comput.*, vol. 17, no. 2, pp. 307–319, 2018.

[15] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "TRAC: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *IEEE INFOCOM*, 2014.

[16] J. Wang, Y. Wang, D. Zhang, F. Wang, H. Xiong, C. Chen, Q. Lv, and Z. Qiu, "Multi-task allocation in mobile crowd sensing with individual task quality assurance," *IEEE. Trans. Mob. Comput.*, vol. 17, no. 9, pp. 2101–2113, 2018.

[17] X. Gao, S. Chen, and G. Chen, "Mab-based reinforced worker selection framework for budgeted spatial crowdsensing," *IEEE Trans. Knowl. Data Eng.*, pp. 1–1, 2020.

[18] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," *Mach. Learn.*, vol. 47, no. 2-3, pp. 235–256, 2002.

[19] J. L. Lihong Li, Wei Chu and R. E. Schapire, "A contextual-bandit approach to personalized news article recommendation," in *ACM WWW*, 2010.

[20] Y. Xia, Q. Tao, W. Ma, N. Yu, and T. Y. Liu, "Budgeted multi-armed bandits with multiple plays," in *IJCAI*, 2016.

[21] R. B. Myerson, "Optimal auction design," *Math. Oper. Res.*, vol. 6, no. 1, pp. 58–73, 1981.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMC.2021.3059346, IEEE Transactions on Mobile Computing

[22] W. Chen, Y. Wang, and Y. Yuan, "Combinatorial multi-armed bandit: general framework, results and applications," in *ACM ICML*, 2013.

[23] S. Yang, F. Wu, S. Tang, T. Luo, X. Gao, L. Kong, and G. Chen, "Selecting most informative contributors with unknown costs for budgeted crowdsensing," in *IEEE IWQoS*, 2016.

[24] Eugene, L., and Lawler, "Fast approximation algorithms for knapsack problems," *Math. Oper. Res.*, vol. 4, no. 4, pp. 339–356, 1979.

[25] V. V. Vazirani, *Approximation algorithms*. Springer Verlag, 2001.

[26] "Chicago taxi trips," https://www.kaggle.com/chicago/chicagotaxi-trips-bq.

[27] Z. Wang, J. Li, J. Hu, J. Ren, Z. Li, and Y. Li, "Towards privacy-preserving incentive for mobile crowdsensing under an untrusted platform," in *IEEE INFOCOM*, 2019.

[28] Z. Wang, X. Pang, Y. Chen, H. Shao, Q. Wang, L. Wu, H. Chen, and H. Qi, "Privacy-preserving crowd-sourced statistical data publishing with an untrusted server," *IEEE. Trans. Mob. Comput.*, vol. 18, no. 6, pp. 1356–1367, 2019.

[29] Z. Zheng, F. Wu, X. Gao, H. Zhu, S. Tang, and G. Chen, "A budget feasible incentive mechanism for weighted coverage maximization in mobile crowdsensing," *IEEE. Trans. Mob. Comput.*, vol. 16, no. 9, pp. 2392–2407, 2017.

[30] Y. Han and H. Wu, "Minimum-cost crowdsourcing with coverage guarantee in mobile opportunistic d2d networks," *IEEE. Trans. Mob. Comput.*, vol. 16, no. 10, pp. 2806–2818, 2017.

[31] L. Zheng and L. Chen, "Maximizing acceptance in rejection-aware spatial crowdsourcing," *IEEE Trans. Knowl. Data Eng.*, vol. 29, no. 9, pp. 1943–1956, 2017.

[32] J. Wang, Y. Wang, L. Wang, and Y. He, "Gp-selector: a generic participant selection framework for mobile crowdsourcing systems," *World Wide Web*, vol. 21, no. 3, pp. 759–782, 2018.

[33] N. Chen, T. Qiu, X. Zhou, K. Li, and M. Atiquzzaman, "An intelligent robust networking mechanism for the internet of things," *IEEE Commun. Mag.*, vol. 57, no. 11, pp. 91–95, 2019.

[34] T. Qiu, J. Liu, W. Si, and D. O. Wu, "Robustness optimization scheme with multi-population co-evolution for scale-free wireless sensor networks," *IEEE/ACM Trans. Netw.*, vol. 27, no. 3, pp. 1028–1042, 2019.

[35] Q. Xu and R. Zheng, "When data acquisition meets data analytics: A distributed active learning framework for optimal budgeted mobile crowdsensing," in *IEEE INFOCOM*, 2017.

[36] T. Qiu, H. Wang, K. Li, H. Ning, A. K. Sangaiah, and B. Chen, "Sigmm: A novel machine learning algorithm for spammer identification in industrial mobile cloud computing," *IEEE Trans. Ind. Informatics*, vol. 15, no. 4, pp. 2349–2359, 2019.

[37] J. Wang, Y. Wang, D. Zhang, F. Wang, Y. He, and L. Ma, "Psallocator: Multi-task allocation for participatory sensing with sensing capability constraints," in *ACM CSCW*, 2017.

[38] G. Bordogna, P. Carrara, L. Criscuolo, M. Pepe, and A. Rampini, "On predicting and improving the quality of volunteer geographic information projects," *Int. J. Digit. Earth*, vol. 9, no. 2, pp. 134–155, 2016.

[39] S. Jain, B. Narayanaswamy, and Y. Narahari, "A multiarmed bandit incentive mechanism for crowdsourcing demand response in Smart Grids," in *AAAI*, 2014.

[40] A. Biswas, S. Jain, D. Mandal, and Y. Narahari, "A truthful budget feasible multi-armed bandit mechanism for crowdsourcing time critical tasks," in *Springer AAMAS*, 2015.

[41] S. Jain, G. Ghalme, S. Bhat, S. Gujar, and Y. Narahari, "A deterministic MAB mechanism for crowdsourcing with logarithmic regret and immediate payments," in *Springer AAMAS*, 2016.

[42] Y. Wu, F. Li, L. Ma, Y. Xie, T. Li, and Y. Wang, "A context-aware multiarmed bandit incentive mechanism for mobile crowd sensing systems," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 7648–7658, 2019.

Mingjun Xiao is a professor in the School of Computer Science and Technology at the University of Science and Technology of China (USTC). He received his Ph.D. degree from USTC in 2004. His research interests include mobile crowdsensing, blockchain, edge computing, mobile social networks, vehicular ad hoc networks, auction theory, data security and privacy. He has published more over 90 papers in referred journals and conferences, including TPDS, TC, TMC, TON, TKDE, TSC, INFOCOM, ICNP, ICDCS, etc. He served as the TPC member of many top conferences, including INFOCOM20, INFOCOM19, ICDCS19, DASFAA19, INFOCOM18, etc. He is on the reviewer board of several top journals such as TMC, TON, TPDS, TSC, TVT, TCC, etc.

Baoyi An is a PhD student in the School of Computer Science and Technology at the University of Science and Technology of China (USTC) currently. She received her BS degree from the School of Computer Science and Engineering at the Nanjing Normal University (NJNU), China, in 2017. Her research interests include data trading, blockchain, privacy preservation, and incentive mechanism.

Jing Wang is a master student in the School of Computer Science and Technology at the University of Science and Technology of China (USTC) currently. She received her BS degree from Xidian University, China, in 2018. Her research interests include crowdsourcing, reinforcement learning, and incentive mechanism.

Guoju Gao is an assistant professor in the School of Computer Science and Technology at the Soochow University. He received his Ph.D. degree in computer science and technology with the School of Computer Science and Technology, University of Science and Technology of China, Hefei, China, in 2020. He visited the Temple University, USA, from Jan. 2019 to Jan. 2020. His research interests include privacy preservation, incentive mechanisms, and reinforcement learning. He has published 20 articles in referred journals and conferences, including IEEE TMC, ToN, TPDS, TSC, IoT Journal, INFOCOM, ICPP, etc.

Sheng Zhang (Member, IEEE) received the B.S. and Ph.D. degrees from Nanjing University in 2008 and 2014, respectively. He is currently an Associate Professor with the Department of Computer Science and Technology, Nanjing University. He is also a member of the State Key Laboratory for Novel Software Technology. To date, he has published more than 70 articles, including those that appeared in TMC, TON, TPDS, TC, MobiHoc, ICDCS, INFOCOM, SECON, IWQoS, and ICPP. His research interests include cloud computing and edge computing. He is a Senior Member of the CCF. He received the Best Paper Runner-Up Award from IEEE MASS 2012. He was a recipient of the 2015 ACM China Doctoral Dissertation Nomination Award.

Wu Jie is the Director of the Center for Networked Computing and Laura H. Carnell professor at Temple University. He also serves as the Director of International Affairs at College of Science and Technology. He served as Chair of Department of Computer and Information Sciences from the summer of 2009 to the summer of 2016 and Associate Vice Provost for International Affairs from the fall of 2015 to the summer of 2017. Prior to joining Temple University, he was a program director at the National Science Foundation and was a distinguished professor at Florida Atlantic University. His current research interests include mobile computing and wireless networks, routing protocols, cloud and green computing, network trust and security, and social network applications. Dr. Wu regularly publishes in scholarly journals, conference proceedings, and books. He serves on several editorial boards, including IEEE Transactions on Mobile Computing, IEEE Transactions on Service Computing, Journal of Parallel and Distributed Computing, and Journal of Computer Science and Technology. Dr. Wu was general co-chair for IEEE MASS 2006, IEEE IPDPS 2008, IEEE ICDCS 2013, ACM MobiHoc 2014, ICPP 2016, and IEEE CNS 2016, as well as program chair/cochair for IEEE INFOCOM 2011, CCF CNCC 2013, and ICCCN 2020. He was an IEEE Computer Society Distinguished Visitor, ACM Distinguished Speaker, and chair for the IEEE Technical Committee on Distributed Processing (TCDP). Dr. Wu is a Fellow of the AAAS and a Fellow of the IEEE. He is the recipient of the 2011 China Computer Federation (CCF) Overseas Outstanding Achievement Award.