A Context-Aware Multi-Armed Bandit Incentive Mechanism for Mobile Crowd Sensing Systems

Yue Wu, Fan Li, *Member, IEEE,* Liran Ma, *Member, IEEE,* Yadong Xie, Ting Li, and Yu Wang, *Fellow, IEEE*

***Abstract*—Smart City is a key component in Internet of Things (IoTs), so it has attracted much attention. The emergence of Mobile Crowd Sensing (MCS) systems enables many smart city applications. In an MCS system, sensing tasks are allocated to a number of mobile users. As a result, the sensing related context of each mobile user plays a significant role on service quality. However, some important sensing context is ignored in the literature. This motivates us to propose a Context-aware Multi-Armed Bandit (C-MAB) incentive mechanism to facilitate quality-based worker selection in an MCS system. We evaluate a worker's service quality by its context (i.e., extrinsic ability and intrinsic ability) and cost. Based on our proposed C-MAB incentive mechanism and quality evaluation design, we develop a Modified Thompson Sampling Worker Selection (MTS-WS) algorithm to select workers in a reinforcement learning manner. MTS-WS is able to choose effective workers because it can maintain accurate worker quality information by updating evaluation parameters according to the status of task accomplishment. We theoretically prove that our C-MAB incentive mechanism is selection efficient, computationally efficient, individually rational, and truthful. Finally, we evaluate our MTS-WS algorithm on simulated and real-world datasets in comparison with some other classic algorithms. Our evaluation results demonstrate that MTS-WS achieves the highest cumulative utility of the requester and social welfare.**

***Index Terms*—participant selection, mobile crowd sensing, multi-armed bandit**

I. INTRODUCTION

Mobile Crowd Sensing (MCS) is emerging as an attractive paradigm for accomplishing complicated tasks by tapping into the power of a large number of contributors (i.e., workers or pariticipants) [1], [2]. Example of these tasks include urban traffic information mapping [3], [4], parking lot searching [5], massive dataset labelling [6], and health care provisioning [7]. Recent advances in mobile computing technologies and Internet of Things (IoTs) technologies enable us to place more built-in sensors and wireless communication modules into a smart device or IoT device. As a result, MCS is expected to inspire more novel applications (especially smart city applications) and greatly enhance the quality of our daily lives in various aspects. However, there still are many challenges that need to be addressed so as to meet this expectation.

One of the fundamental challenges for an MCS system is the design of its incentive mechanism. An effective incentive mechanism needs to enable the crowd sensing requester to select high-quality workers with reasonable expenses in order to complete sensing tasks successfully. Thus, worker selection becomes one of key issues in designing an incentive mechanism. Many existing works formulate the worker selection problem as an optimization problem, and attempt to strike a balance among constraints such as sensing cost, task coverage, energy efficiency, or user privacy [8]–[13]. Under these constraints, workers are selected based on their context information (i.e., extrinsic ability) that typically includes location, buffer size, and battery power [14]–[17]. However, workers' personal characteristics (i.e., intrinsic ability) such as qualifications and interests are often ignored in the worker selection process. These characteristics can also have considerable impacts on service quality. Therefore, it is of great significance to extend the concept of context information to include both extrinsic and intrinsic abilities into worker selection consideration.

Since we propose workers' extrinsic and intrinsic abilities in the selection process, we need to design a novel context-aware selection algorithm based on our new worker ability model. Our algorithm design is inspired by the Multi-Armed Bandit (MAB) problem, which refers to the scenario where a gambler is faced with a number of slot machines (i.e., multi-armed bandits). Each bandit provides a random reward from a probability distribution specific to that machine. The gambler's goal is to maximize the sum of rewards earned through a sequence of selected lever pulls. Initially, the gambler has to explore, and thus, may choose seemingly suboptimal arms to gather information. The gathered information is exploited to refine the expected rewards of these arms' and in turn improve long-term rewards [18]. The fundamental challenge in MAB is how to strike a balance between exploration and exploitation.

The MAB problem is similar to our worker selection problem in that the requester (like the gambler) also needs to choose a number of high-quality workers (like the bandits) in a certain sequence so as to successfully complete sensing tasks under budgetary constraints. Traditionally, *Thompson Sampling* (TS) has been used to solve the MAB problem. However, classical TS [19] assumes each arm follows a distribution with a fixed parameter. Nonetheless, a worker's context information (especially its extrinsic ability) changes over time in an MCS

Y. Wu and F. Li and Y. Xie are with the School of Computer Science, Beijing Institute of Technology, Beijing Engineering Research Center for High Volume Language Information Processing and Cloud Computing Applications, Beijing, 100081, China (e-mail: ywu@bit.edu.cn, fli@bit.edu.cn, ydxie@bit.edu.cn).

L. Ma is with Department of Computer Science, Texas Christian University, Fort Worth, TX 76129, USA (e-mail: L.MA@tcu.edu).

T. Li and Y. Wang are with Department of Computer Science, University of North Carolina at Charlotte, Charlotte, NC 28223, USA (e-mail: tli8@uncc.edu, yu.wang@uncc.edu).

Manuscript received September 29, 2018; revised December 17, 2018;accepted Febuary 18, 2019.

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/JIOT.2019.2903197, IEEE Internet of Things Journal

system, for which classical TS cannot be applied directly in our scenario. This presents another challenge that how to update the evaluations of workers for the worker selection problem. Reinforcement learning is one of the most active research areas in artificial intelligence. This computational approach can learn whereby an agent tries to maximize the total amount of reward it receives when interacting with a complex, uncertain environment. Thus, reinforcement learning methods have the advantages of handling highly dynamic context of workers.

Motivated by the aforementioned challenges, we propose an effective Context-aware MAB (C-MAB) incentive mechanism that enables the requester to choose a group of workers with high quality-price ratio to execute sensing tasks, and guarantees the utility and truthfulness of the workers. In our proposed C-MAB mechanism, workers' service quality are estimated by combining their extrinsic ability and intrinsic abilities. Next, we design an online worker selection algorithm, termed as Modified Thompson Sampling Worker Selection (MTS-WS), which combines classical TS with reinforcement learning. MTS-WS selects workers by updating the evaluation parameters of workers according to their status of accomplishment. The main contributions of this paper are as follows:

- We take workers' extrinsic and intrinsic abilities into worker selection consideration, and update the evaluation parameters of workers according to their status of accomplishment in a reinforcement learning way.
- We investigate the relationships between these two types of abilities and workers' service quality, and build a novel worker ability model.
- We formulate the worker selection problem as a novel MAB problem and propose a context-aware MAB incentive mechanism.
- We design an online MTS-WS algorithm for our C-MAB incentive mechanism. MTS-WS selects workers based on their service quality effectively in a reinforcement learning manner. We theoretically prove that MTS-WS possesses many desirable properties, including selection efficiency, computational efficiency, individual rationality, as well as truthfulness.

The remainder of the paper is organized as follows. We present related work in Section II. In Section III and Section IV, we introduce our system model and formulate the online worker selection problem. In Section V, we propose our algorithm based on Thompson Sampling and theoretically prove that our mechanism is selection efficient, individually rational, truthful and computationally efficient. In Section VI, we study features and present our performance evaluation results. Finally, we draw our conclusion in Section VII.

II. RELATED WORK

In this section, we present our investigation on worker selection problems in MCS. Since our incentive mechanism is based on MAB model, we also investigate MAB algorithms.

A. Worker Selection in MCS

The popularity of online MCS systems drove the emergence of frameworks, platforms and incentive mechanisms [20]–[24]. Similar to many existing studies, TaskMe [25] and CrowdTasker [10] were built on a central service entity, which was called a platform, to collect information of both requesters and workers. The platform chose suitable workers and assigned tasks to achieve an optimal global utility. In our C-MAB incentive mechanism, a requester can contact workers directly through cellular or Wi-Fi connections without such a platform. This self-organized fashion can provide a rapid response policy that requesters proactively recruit workers for newly created tasks instead of posting the tasks online and passively wait for the workers to participate.

Recently, many researchers paid attention to service quality control in worker selection. Jin *et al.* [26] combined quality information with a reverse combinatorial auction in their incentive mechanism design. Wen *et al.* [27] proposed a quality-driven auction based incentive mechanism, where the worker was paid off based on the quality of sensed data instead of working time. However, they did not give specifics on how to measure work quality. To address this problem, Liu *et al.* [28] designed a context-aware data quality estimation scheme with a context-quality classifier to guide user recruitment. Another quality estimation measure was proposed based on a two-dimensional trust model in [29]. However, neither of these two works considered updating quality estimation according to task accomplishment status.

Our proposed C-MAB incentive mechanism differs from the existing designs in two main aspects: i) C-MAB extends the context information to include both extrinsic and intrinsic abilities, and employs them to estimate a worker's service quality; and, ii) C-MAB constantly updates the evaluation parameters of a worker's service quality of according to the status of task accomplishment.

B. MAB Algorithms

The fundamental challenge in multi-armed bandit problems is the need for balancing exploration and exploitation. The context-free K-armed bandit problem has been studied by statisticians for a long time [30]–[32]. One of the simplest and most straightforward algorithms was ϵ-greedy [33]. In each time round, ϵ-greedy first estimated the average payoff of each arm according to the historical data. Then, it selected the arm who had the highest payoff estimate with probability $1 - \epsilon$, and selected another arm with ϵ. In contrast to the unguided exploration strategy adopted by ϵ-greedy, another class of algorithm was generally known as Upper Confidence Bound (UCB) algorithm [34]–[36], which estimated the mean payoff of each arm and corresponding confidence interval. Subsequently, the arm that achieved the highest upper confidence bound was selected. Although context-free K-armed bandits were extensively studied, the more general contextual bandit problem remained unsolved. The EXP4 algorithm [37] used an exponential weighting technique to achieve considerable rewards, but the computational complexity may grow exponentially with increasing number of features. Another general contextual bandit algorithm was the epoch-greedy algorithm [38], which was similar to ϵ-greedy with shrinking ϵ. This algorithm was computationally efficient given an oracle

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/JIOT.2019.2903197, IEEE Internet of Things Journal

optimizer, but had a weaker regret guarantee. Compared to ϵ-greedy, UCB and EXP4, our MTS-WS algorithm can achieve both computational efficiency and bounded regret.

Combinatorial multi-armed bandit problem also has become an active research area [39], [40], where simple arms with unknown distributions form super arms. In each selection round, a super arm is played and the outcomes are observed, which helps the selections of super arms in future rounds. There are also some modified algorithms based on Thompson Sampling designed for this problem [41].Different from combinatorial multi-armed bandit problem, in crowd sensing scenario, selection mechanisms should focus on each participant rather than combinations of participants. It is unnecessary to explore all combinations of participants. In addition, participants are all independent from each other, which means an execution result of a participant does not affect others' execution results. Thus, it is also unnecessary to share the execution results of one group with another group. Our proposed C-MAB mechanism can fit crowd sensing scenarios better.

Context-aware multi-armed bandit problem is another research point and several related models have been applied into crowd sensing systems. [42] defined a framework for online spatial task assignment based on a contextual bandit algorithm. [43] proposed an algorithm called bounded ε-first based on multi-armed bandit model to promote efficiency of crowdsourcing systems. [44] addressed the task assignment in crowdsourcing by the proposed bandit-based task assignment method with a least confidence strategy. Compared to these works, C-MAB has a stronger scalability for heterogeneous sensing tasks as requesters can define favourable assessment parameters of participants' expertise.

III. SYSTEM MODEL

The MCS system model considered in this paper consists of a requester and n workers (represented by $W = \{w_1, w_2, \ldots, w_n\}$). Workers can execute a set of sensing tasks denoted as $\mathcal{T} = \{\tau_1, \tau_2, \ldots\}$. We assume that the requester publishes tasks and collects results in a self-organized manner. Lastly, the requester contacts workers through cellular or Wi-Fi connections.

A. Overall Flow

At first, the requester publishes a sensing task and sends it to surrounding mobile users. Then some mobile users send their context information and bids to the requester and take part in the task as workers. After the requester estimates the service quality of all workers, one of the three following cases may happen:

Case 1: If the requester finds a worker (e.g., Worker1 in Fig. 1) that is not suitable for a task, it does nothing on the worker.

Case 2: If the requester finds a worker (e.g., Worker2 in Fig. 1) that is suitable for a task, it sends an allocation request to the worker. If the worker fails to finish the task, it cannot get a payment.

Case 3: If the requester finds a worker (e.g., Worker3 in Fig. 1) that is suitable for a task, it sends an allocation request to the worker. When the worker finishes the task successfully and submits the result, the requester rewards a payment to the worker.

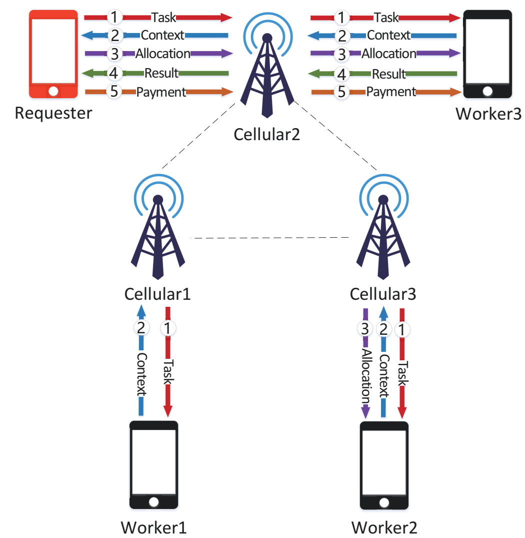

Fig. 1: Three cases of the interactions between a requester and a worker.

If the requester does not collect enough results, it repeats the worker selection procedure until enough results are collected or the sensing task expires.

B. Task Model

In this paper, a sensing task is associated with a set of attributes including *Type, Description, Amount*, and *maxTime*.

- Type: The kind of the task such as information collecting or target finding.
- Description: A brief description of the task in a textual form.
- Amount: The number of results that the requester needs for the task.
- MaxTime: The maximum allowed time within which the requester needs to collect the required amount of results.

Once a requester decides to launch a sensing task, it sends information of these attributes to surrounding mobile users.

C. Worker Ability Model

A worker's performance is mainly determined by two types of abilities: i) extrinsic ability, and ii) intrinsic ability. The former is decided by a worker's current situation such as location, moving speed, the distance between the worker and task, and the remaining battery power of the sensing device. Typically, the extrinsic ability is called dynamic context as it changes with time. The latter is decided by a worker's personal characteristics such as its qualifications and interests for certain type of tasks. Contrary to the extrinsic ability, the intrinsic ability is called static context because it reflects personal qualifications and traits which can always remain steady for a period of time.

Note that it is important that the requester chooses a worker by examining both dynamic and static context information according to the attributes of the sensing task. For example, the requester publishes a target finding task to a number of mobile users. A worker that is close to the target may be a good candidate. However, if the worker is not interested in a target finding task, it may lead to the failure of this task.

We denote the dynamic context of worker w_i at time t by a $1 \times K$ vector $C_i(t) = [c_{i_1}(t), c_{i_2}(t), \ldots, c_{i_K}(t)]$, where c_{i_K} denotes the K-th feature. As different features have different effects on tasks, we formulate the effect as a $K \times 1$ vector $Z_j = [z_{j_1}, z_{j_2}, \ldots, z_{j_K}]^T$, where z_{j_K} is the parameter to measure the importance of the K-th feature in task τ_j. Therefore, the extrinsic ability of a worker w_i with respect to task τ_j at time t is given as follows:

$$q_{i,j}(t) = C_i(t) \cdot Z_j = \sum_{k=1}^{K} c_{i_k}(t) \cdot z_{j_k}. \tag{1}$$

It is worth mentioning that this general model fits a variety of MCS tasks. For example, for a location-aware sensing task, the distance between the worker and task location is more important than other features. While, for a photo or voice collecting task, the storage size of the worker's sensing device is more important, as it allows to take high-definition pictures or sound tracks. We will explore how to set the optimal feature parameters according to a task type in our future work.

Since a worker's intrinsic ability is determined by its personal characteristics, it is difficult to measure directly. Instead, we use other information to estimate the intrinsic ability. Specifically, we denote $r_{i,j} \in \{0,1\}$ as the execution results of w_i. If w_i finishes the task τ_j successfully then $r_{i,j} = 1$, otherwise $r_{i,j} = 0$. In this paper, we focus on fully completions of tasks, as most of mobile computing tasks are relatively simple or highly integrated that cannot be partitioned. For example, for a target finding task, the execution result only can be "find" or "not find". And for a noise collecting task, the execution result only can be "have collected" or "can not collect". For a certain sensing task, a participant has a probability of finishing it because of various factors. Therefore, we suppose $r_{i,j}$ follows a Bernoulli distribution $B(\mu_i)$ with μ_i [19]. We can use the mean μ_i of $r_{i,j}$ to reflect the intrinsic ability of w_i.

D. Utility Model

Assume that a sensing task τ_j needs m pieces of results. The requester sets v_j as the value of each piece. Once worker w_i finishes a task successfully, it gets a payment $p_{i,j}$. Then the utility of the requester $U_{requester}$ is defined as follows

$$U_{requester} = mv_j - \sum_{i=1}^{n} r_{i,j} p_{i,j}. \tag{2}$$

Generally, a worker's cost is the same as its bid b_i. Thus, the utility of each worker U_{w_i} is defined as

$$U_{w_i} = r_{i,j}(p_{i,j} - b_{i,j}). \tag{3}$$

Subsequently, the social welfare U_{social} can be defined as

$$\begin{aligned} U_{social} &= U_{requester} + \sum_{i=1}^{n} U_{w_i} \\ &= mv_j - \sum_{i=1}^{n} r_{i,j} b_{i,j}. \end{aligned} \tag{4}$$

Since we consider the problem for an arbitrary task, task index j is omitted in all variables for simplicity in the following sections. For example, $q_i(t)$, b_i, r_i and p_i is short for $q_{i,j}(t)$, $b_{i,j}$, $r_{i,j}$ and $p_{i,j}$. Key notations used in this paper are listed in Table I.

TABLE I: notation used

Notation	Description
n	Number of workers
m	Number of required results of task
v	Value of each copy of result
w_i, τ_j	i-th worker and j-th task
$C_i(t)$	Context vector of w_i at time t
Z_j	Importance vector of context under task τ_j
$q_i(t)$	Extrinsic ability of w_i at time t
r_i	Execution result of w_i at time t
μ_i	The mean of r_i
α_i, β_i	Parameters of Beta distribution
$b_i, \hat{b}_i$	Actual and dummy bid of w_i
$p_i, \hat{p}_i$	Payment of bid b_i and $\hat{b}_i$ by w_i
$Q_i(t)$	Service quality of w_i at time t
W	Worker set
W_s	Winning worker set
W_*	Optimal worker set
h_i	The historical record of w_i
B, P, H	The set of all bids, payments and history
T	The time limitation of a task

IV. PROBLEM FORMULATION

In this section, we formally introduce the online worker selection problem and discuss the design objectives of our proposed incentive mechanism.

A. Online Worker Selection Problem

On one hand, we observe that based on specific context, workers with high-quality service should be selected with high probability through an online evaluation. On the other hand, the extrinsic ability of each worker varies significantly over time. Based on these two reasons, we formulate the online worker selection problem as a context-aware MAB problem. In the process of worker selection, the requester chooses a worker strategically which is comparable to pulling the lever of an armed bandit. The execution results provided by the chosen worker is regarded as the reward of the chosen bandit.

In a MAB problem, the player wants to maximize cumulative rewards. Similarly, the requester in MCS systems wants to maximize profits while cutting down expenses. As μ_i indicates the intrinsic ability of w_i and belongs to $(0,1)$, we use $\mu_i q_i(t)$ to represent the working ability of w_i. Then we define a service quality concept by jointly taking a worker's extrinsic ability, intrinsic ability and its bid into consideration as

$$Q_i(t) = \frac{\mu_i q_i(t)}{b_i}. \tag{5}$$

Therefore, our policy for online worker selection problem is to find the optimal worker set $\mathcal{W}$ which maximizes the sum of workers' service qualities, i.e.,

$$\mathcal{W} \triangleq \underset{W_s}{\arg\max} \sum_{l=1}^{m} Q_{s_l}(t_{s_l}), \tag{6}$$

where $W_s = \{w_{s_1}, w_{s_2}, \ldots, w_{s_m}\}$ is the winning worker set constructed by our MTS-WS algorithm. t_{s_l} represents the time when worker w_{s_l} is chosen and $Q_{s_l}(t_{s_l})$ represents the service quality of w_{s_l}.

B. Design Objectives

In this paper, we aim to ensure that our worker selection policy has the following advantageous properties.

We denote $W_* = \{w_{*_1}, w_{*_2}, \ldots, w_{*_m}\}$ as the optimal worker set. Both W_* and W_s are multiset and have m elements. Then we can define the regret as the difference of service quality between W_* and W_s.

$$regret = \sum_{u=1}^{m} \left(\frac{\mu_{*_u} q_{*_u}}{b_{*_u}} - \frac{\mu_{s_u} q_{s_u}}{b_{s_u}}\right). \tag{7}$$

Definition 1. *(Selection efficiency): Mechanism M is selection efficient if its regret is bounded.*

Definition 2. *(Computational efficiency): Mechanism M is computationally efficient if it can be executed within polynomial time.*

Definition 3. *(Individual Rationality): Mechanism M is individually rational if for each worker, its utility is non-negative when reporting its true bid. Formally,*

$$U_{w_i} \geq 0, \quad \forall w_i \in W. \tag{8}$$

Definition 4. *(Truthfulness): Mechanism M is truthful if and only if bidding b_i is the dominant strategy for each worker w_i. Formally*

$$p_i - b_i \geq \hat{p}_i - \hat{b}_i, \quad \forall w_i \in W, \tag{9}$$

where $\hat{b}_i$ is a dummy bid of w_i which is not equal to its cost. And $\hat{p}_i$ is the payment if a worker claims $\hat{b}_i$.

V. C-MAB INCENTIVE MECHANISM

In this section, we present a worker selection algorithm based on Thompson Sampling for our crowd sensing model. We also prove that our algorithm is selection efficient, individually rational, computationally efficient, and truthful.

A. MTS-WS Algorithm

1) Probabilistic Model: Our work selection policy attempts to find a group of mobile users with high service qualities. In previous sections, we assume that r_i follows a Bernoulli distribution $B(\mu_i)$. In this distribution, μ_i is a fixed but unknown parameter for worker w_i. In order to maximize workers' service qualities, we need to predict μ_i at first. According to [19], we assume μ_i follows a Beta distribution $Beta(\alpha_i + 1, \beta_i + 1)$. $Beta(\alpha_i + 1, \beta_i + 1)$ is the prior distribution of μ_i which indicates the subjective cognition of μ_i under historical information or common sense. For example, if a worker w_i has executed sensing tasks for 10 times and there are 6 times finished and 4 times failed, the mean of r_i respected to the 10 times execution is 0.6. Then we can conjecture the real value of μ_i approaches 0.6.

Based on the definition of prior distribution, we can calculate α_i and β_i according to historical execution information of w_i. Specifically, α_i is the frequency that the worker w_i finishes the task successfully and β_i is the frequency that the worker w_i fails. Therefore, once we get an execution result r_i of w_i, we can update α_i and β_i by the following formulae:

$$\begin{aligned} \alpha_i &\leftarrow \alpha_i + 1, \quad when \quad r_i = 1; \\ \beta_i &\leftarrow \beta_i + 1, \quad when \quad r_i = 0. \end{aligned} \tag{10}$$

For convenience, we suppose each worker w_i maintains a historical record h_i which contains μ_i, α_i and β_i. Note that for workers without historical results, we initialize $\alpha_i = 0$ and $\beta_i = 0$.

Now it is easy to understand the basic idea of our worker selection algorithm based on TS. For each worker, first we get α_i and β_i to construct the prior distribution $Beta(\alpha_i + 1, \beta_i + 1)$ according to historical records. Then we sample μ_i from the prior distribution to calculate $Q_i(t)$. After we select workers whose $Q_i(t)$ is high and observe their execution r_i, their α_i and β_i are updated for next usage. This reinforcement learning manner maintains a balance between the exploration and exploitation of μ_i.

$$Pr(r_i = x) = Pr(r_i = x|\mu_i) \cdot Pr(\mu_i|\alpha_i, \beta_i). \tag{11}$$

Equation (11) demonstrates how to calculate the probability of $r_i = x (x \in \{0, 1\})$. Combined equation (10) and equation (11), we can see that MTS-WS is reasonable. If $r_i = 0$, the mean of μ_i which is $\alpha_i/(\alpha_i + \beta_i)$ decreases. Then the probability that μ_i samples a large value decreases. In another word, the probability that w_i is chosen next time decreases. If $r_i = 1$, the mean of μ_i increases. Then the probability that w_i is chosen next time increases. No matter $r_i = 0$ or $r_i = 1$, the variance of μ_i which is $\dfrac{\alpha_i \beta_i}{(\alpha_i + \beta_i)^2 (\alpha_i + \beta_i + 1)}$ decreases. This corresponds to the fact that as the number of experiments increases, the variance decreases.

Algorithm 1: Preprocessing

input W, H;
for all w_i in W **do**
    if $b_i > v$ **then**
        $W \leftarrow W \backslash w_i$;
    end if
end for
for all w_i in W **do**
    if h_i is $\emptyset$ **then**
        $\alpha_i \leftarrow 0$, $\beta_i \leftarrow 0$;
    end if
end for
return W;

Algorithm 2: Worker Selection

1: input W, H, B;
2: $t \leftarrow 0$;
3: $s \leftarrow m$;
4: **while** $s \neq 0$ **and** $t < T$ **do**
5:   **for all** w_i in W **do**
6:     sample $\mu_i \sim Beta(\alpha_i + 1, \beta_i + 1)$;
7:     $Q_i(t) \leftarrow \frac{\mu_i q_i(t)}{b_i}$;
8:   **end for**
9:   $W' \leftarrow W$;
10:   **for** $j \leftarrow 1$ to s **do**
11:     $w'_s \leftarrow \arg\max_{w_i \in W'} \; Q_i(t)$;
12:     $W' \leftarrow W' \backslash w'_s$;
13:     $W'_s \leftarrow W'_s \cup w'_s$;
14:   **end for**
15:   observe results;
16:   **for all** w_i in W'_s **do**
17:     **if** $r_i = 1$ **then**
18:       $W_s \leftarrow W_s \uplus w_i$;
19:       $s \leftarrow s - 1$;
20:       $\alpha_i \leftarrow \alpha_i + 1$;
21:     **end if**
22:     **if** $r_i = 0$ **then**
23:       $\beta_i \leftarrow \beta_i + 1$;
24:     **end if**
25:   **end for**
26:   Pricing;
27:   $t \leftarrow t + 1$;
28: **end while**
29: record α_i and β_i into h_i;
30: return W_s, P, H;

2) Complete Algorithm: Now we present MTS-WS in details. Algorithm 1 is the preprocessing of MTS-WS. We first remove those workers whose bid is higher than the task value (Line $3-4$). Then we check the historical records of each worker. If a worker w_i has no historical record, we initialize α_i, and β_i for it (Line $8-9$).

After preprocessing, we select workers according to Algorithm 2. We first sample μ_i from $Beta(\alpha_i + 1, \beta_i + 1)$ and calculate $Q_i(t)$ for each w_i (Line $6-7$). Then we sort workers by $Q_i(t)$ in a non-increasing order and select the top s workers (Line $11-13$). These workers execute sensing tasks and return results (Line 15). If a worker finish the task successfully, it becomes a winner and is added into a multiset W_s (Line $17-18$). Then the two parameters α_i, and β_i of these s workers are updated and recorded into h_i (Line $20, 23, 29$). If there are some workers who do not finish the task, we select workers again until we collect enough copies of task results or run out of the valid time (Line 4). At last we obtain a winning worker set W_s or the task fails. The time complexity for algorithm 2 is $O(nT)$. We adopt the main idea of reinforcement learning in the process. Reinforcement learning regards learning as a process of tentative evaluation. The agent chooses an action for the environment and the environment changes state after accepting the action. At the same time, an enhanced signal (a reward or punishment) is sent to the agent. If an certain action of the agent leads to a positive reward, then the trend of generating this action will be strengthened. In our C-MAB mechanism, the requester is similar to the agent in reinforcement learning, and selecting an arbitrary participant is like an action. When the requester selects a participants, he takes an action. If the selected participant successfully finishes a sensing task, the requester gets a sensing result as a reward, then the μ_i of this participant increasees which means the trend of selecting this participant is strengthened. The selection action not only affects the immediate enhancement value, but also affects the selection at next moment and the final result.

After obtaining the winning workers in one selection round of Algorithm 2, we need to determine their payments. In algorithm 3, we first take the maximum service quality among the remaining worker set W' as a reference value $Q_{re}(t)$ (Line 2). Then we set $\frac{\mu_i q_i(t)}{Q_i(t)}$ as the payment for w_i if $\frac{\mu_i q_i(t)}{Q_i(t)}$ is smaller than the task value v (Line $4-5$). Else we set v as its payment (Line 7). This pricing policy can guarantee each worker's individual rationality.

B. Analysis of MTS-WS

We now analyze the performance of our proposed MTS-WS.

Lemma 1. *MTS-WS is selection efficient.*

Proof. Let $\mu_{*_i} = \max \mu_i$ and $\Delta_i = \mu^* - \mu_i$. $k_i(t)$ denotes the number of times that the worker w_i has been chosen before time t . $\frac{q}{b}$ is a constant which can be calculated by approximate methods. Then according to the study of Agrawal *et al.* [19], the expected total regret in time T is given by

$$\begin{aligned} E[\sum_{t=1}^{T} regret] &= E[\sum_{t=1}^{T}(\sum_{u=1}^{m} \mu_{*_u} \cdot \frac{q_{*_u}}{b_{*_u}} - \mu_{s_u} \cdot \frac{q_{s_u}}{b_{s_u}})] \\ &\leq \frac{q}{b} E[\sum_{t=1}^{T}(\sum_{u=1}^{m}(\mu_{*_u} - \mu_{s_u}))] \\ &= \frac{q}{b} \cdot \sum_{i=2}^{mn} \Delta_i \cdot E[\sum_{t=1}^{T} k_i(t)] \\ &\leq O(\sum_{i=2}^{mn} \frac{1}{\Delta_i^2}^2 \ln T). \end{aligned} \tag{12}$$

From formula (12), we can conclude that when time T increases, cumulative regret's growth rate becomes lower and lower (because the derivative of $\ln x$ is $\frac{1}{x}$), which reflects the learning procedure of Thompson sampling. $\square$

Lemma 2. *MTS-WS is computationally efficient.*

Proof. Line $5-7$ in algorithm 2 samples μ_i for each work with a loop. At each t, we select workers that are first s-th largest with respect to $Q_i(t)$. This procedure can be implemented by well known linear time selection algorithm whose time complexity is $O(n)$. After observing selected workers' results, we update parameters for each worker according to the result. The time restraint $t < T$ ensures that our algorithm would not

Algorithm 3: Pricing

input W_s, W';
$Q_{re}(t) \leftarrow \arg\max_{w_i \in W'} Q_i(t)$
for all w_{s_i} in W_s **do**
  if $\frac{\mu_i q_i(t)}{Q_{re}(t)} > v$ **then**
    $p_{s_i} \leftarrow v$;
  else
    $p_{s_i} \leftarrow \frac{\mu_i q_i(t)}{Q_{re}(t)}$;
  end if
end for
return P;

be trapped in endless loop. As a result, the selecting step is executed no more than T steps. Therefore, the time complexity for MTS-WS is $O(nT)$. □

Lemma 3. *MTS-WS is individually rational.*

Proof. In the pricing step, we first calculate the maximum service quality in the remaining worker set W' as a reference value $Q_{re}(t)$. Then each winning worker's payment is equal to $\frac{\mu_i q_i(t)}{Q_{re}(t)}$ or v. Since $Q_i(t) > Q_{re}(t)$ and $b_i \leq v$, the utility of worker i satisfies that

$$\begin{aligned} U_{w_i} &= p_i - b_i = \frac{\mu_i q_i(t)}{Q_{re}(t)} - b_i \\ &= \frac{1}{b_i} \cdot \left(\frac{Q_i(t)}{Q_{re}(t)} - 1\right) > 0 \end{aligned} \tag{13}$$

or

$$U_{w_i} = p_i - b_i = v - b_i > 0. \tag{14}$$

□

Lemma 4. *MTS-WS is truthful.*

Proof. A worker may attempt to get a higher utility by lying about its bid. We denote $\hat{b}_i$ as a dummy bid of w_i, and $\hat{q}_i$ as the payment if w_i claim $\hat{b}_i$. Note that $\hat{b}_i$ can be higher or lower than b_i. We assume the winning worker set is $W_s = \{w_{s_1}, w_{s_2}, \ldots, w_{s_m}\}$, and the winning workers are sorted by their service quality Q_{s_i} (we omit t in the proof for convenience) in an decreasing order. We denote w_{m-1} as the first worker whose service quality is smaller than w_{s_m}.

Case 1: If w_i cannot be chosen with b_i, which means

$$\frac{\mu_i q_i}{b_i} < \frac{\mu_{s_m} q_{s_m}}{b_{s_m}},$$

it changes its bid and there will exist three situations:

Case 1.1: If w_i increases its bid ($\hat{b}_i > b_i$), then its service quality will become lower and it also cannot be chosen, thus its utility will maintain zero.

Case 1.2: If w_i decreases its bid, which makes

$$\frac{\mu_i q_i}{\mu_{s_m} q_{s_m} b_{s_m}} < \hat{b}_i < b_i,$$

the service quality of w_i will be smaller than the service quality of w_{s_m}. Then w_i also cannot be chosen and maintains zero utility.

Case 1.3: If w_i decreases its bid, which makes

$$\hat{b}_i > \frac{\mu_i q_i}{\mu_{s_m} q_{s_m} b_{s_m}},$$

the service quality of w_i will be bigger than the service quality of w_{s_m}. Then w_i can be chosen. Its payment satisfies

$$\hat{p}_i = \frac{\mu_i q_i b_{sm}}{\mu_{sm} q_{sm}}.$$

However,

$$\frac{\mu_i q_i}{b_i} < \frac{\mu_{s_m} q_{s_m}}{b_{s_m}}.$$

Thus its utility is

$$U_{w_i} = \hat{p}_i - b_i = \frac{\mu_i q_i b_{sm}}{\mu_{sm} q_{sm}} - b_i < 0.$$

Case 2: If w_i can be chosen with b_i, which means

$$\frac{\mu_i q_i}{b_i} > \frac{\mu_{s_m} q_{s_m}}{b_{s_m}},$$

it changes its bid and there will exist three situations:

Case 2.1: If w_i decreases its bid ($\hat{b}_i < b_i$), then w_i also can be chosen. But according to the pricing algorithm, when μ_i and q_i of w_i do not change, its payment satisfies

$$\hat{q}_i = \frac{\mu_i q_i b_{m-1}}{\mu_{m-1} q_{m-1}} = q_i.$$

Thus the utility of w_i will not change.

Case 2.2: If w_i increases its bid, which makes

$$\hat{b}_i > \frac{\mu_i q_i}{\mu_{s_m} q_{s_m} b_{s_m}},$$

the service quality of w_i will be smaller than the service quality of w_{s_m}. Then w_i cannot be chosen with $\hat{b}_i$. Thus its utility will become zero.

Case 2.3: If w_i increases its bid, which makes

$$b_i < \hat{b}_i < \frac{\mu_i q_i}{\mu_{s_m} q_{s_m} b_{s_m}},$$

the service quality of w_i will be bigger than the service quality of w_{s_m}. Then w_i can be chosen and its payment is

$$\hat{q}_i = \frac{\mu_i q_i b_{m-1}}{\mu_{m-1} q_{m-1}} = q_i.$$

Thus the utility of w_i will not change. □

In summary, based on the above lemmas, we can conclude the following theorem.

Theorem 1. *MTS-WS is selection efficient, computationally efficient, individually rational and truthful.*

VI. PERFORMANCE EVALUATION

In this section, we provide a detailed description of our experimental configuration and results. We first introduce the three baseline methods for the evaluation of MTS-WS. Then, we explain the evaluation setup and results on simulated data and MIT Reality data [45].

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/JIOT.2019.2903197, IEEE Internet of Things Journal

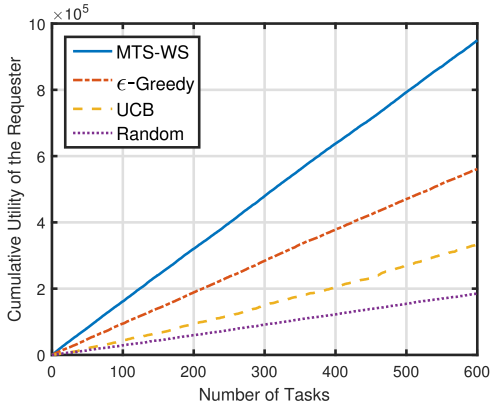

Fig. 2: Comparison of requester's cumulative utility on simulated data.

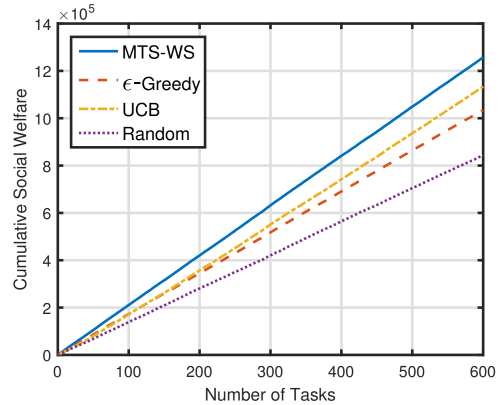

Fig. 3: Comparison of cumulative social welfare on simulated data.

A. Baseline Methods

The three baseline methods are *ϵ-greedy, Upper Confidence Bound (UCB)*, and *Random Selection*. The first two baseline methods are designed for the classical MAB problem. As a result, we need to make minor modifications to fit our settings.

- **ϵ-greedy**: Its original design selects the worker who has the highest service quality with the probability of $1-\epsilon$ and selects another worker randomly with the probability of ϵ. In our evaluation study, we sort workers by $Q_i(t)$ in an increasing order, and select the top $m(1-\epsilon)$ workers. Then, we randomly select $m\epsilon$ workers.
- **UCB**: It selects workers based on the upper confidence bound of the rewards. The original UCB selects the worker whose $\frac{\alpha_i}{\alpha_i+\beta_i}\cdot\frac{q_i}{b_i}+d\sqrt{\frac{q_i(t)^2}{\alpha_i+\beta_i}}$ is the largest. Here, we sort workers by $\frac{\alpha_i}{\alpha_i+\beta_i}\cdot\frac{q_i}{b_i}+d\sqrt{\frac{q_i(t)^2}{\alpha_i+\beta_i}}$ in an increasing order, and select the top m workers.
- **Random Selection**: It selects workers randomly.

B. Evaluation on Simulated Data

We first carry out the evaluation study on simulated data. We generate 600 tasks and their values range from 130 to 150. There are 100 workers in total ($n = 100$) whose bids range from 10 to 140. Each worker's extrinsic ability is in the range of 1 to 15. Each task needs 20 pieces of results ($m = 20$), and maxTime $T = 10$. We run evaluation under the four different algorithms. Then, we record the cumulative utility of the requester and the cumulative social welfare for each algorithm.

Fig. 2 shows the results in terms of the cumulative utility of the requester. As the number of tasks increases, the cumulative utilities of the requester increase under all four algorithms. This means the requester can benefit from workers' participation in MCS. Note that MTS-WS achieves the highest cumulative utility of the requester among all four algorithms.

Fig. 3 shows the cumulative social welfare. Typically, social welfare can be regarded as the income for all participants in our C-MAB model. From the figure, we can see that the growth trends of social welfare are similar to those of the requester's utility. MTS-WS also achieves the highest cumulative social welfare compared with the other three algorithms.

C. Evaluation on MIT Reality Data

Our real-world dataset comes from MIT Reality [45]. It includes data of 100 mobile users from September 2004 to June 2005, including communication records, cell tower ID, Bluetooth discovery records of participants and etc. We select 80 mobile users among the participants and use the data between 09/01/2004 and 12/19/2004 for our evaluation study. We randomly generate several sensing tasks, including Amount and MaxTime. We also randomly generate bids for all workers. If a worker successfully make a call within MaxTime, we consider he/she finishes the task successfully. So the probability of making calls successfully during a period of time is used to be μ_i of the corresponding worker. The communication records, cellular tower IDs and Bluetooth discovery records in MIT Reality are used to construct a worker's dynamic context. The dynamic context includes the following features:

TABLE II: MTS-WS with different context

algorithm name	features combination
TS-WS-1	centrality + activity scope
TS-WS-2	centrality + encounter level
TS-WS-3	activity scope + encounter level
TS-WS-4	centrality + activity scope + encounter level

- Centrality: we use centrality to measure the influence of a worker on the social network it is in. Here, we use the number of different contact IDs (representing different users) of w_i during a selection round as its centrality. We assume that the more contacts of a mobile user has, the greater influence it has.
- Activity scope: A user has the potential to receive more sensing tasks if it can move within a larger activity scope. In MIT's dataset, when a participant is within the coverage of a cellular tower, it is connected and can be localized. Thus, we use the number of different cellular

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/JIOT.2019.2903197, IEEE Internet of Things Journal

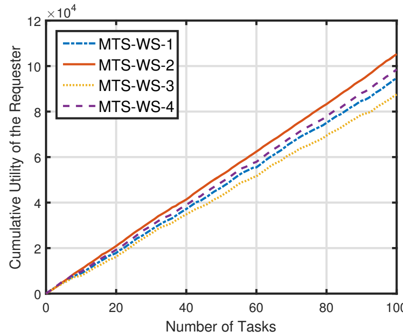

Fig. 4: Comparison of requester's cumulative utility on simulated data.

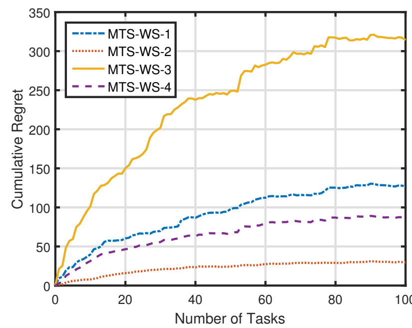

Fig. 5: Comparison of cumulative social welfare on simulated data.

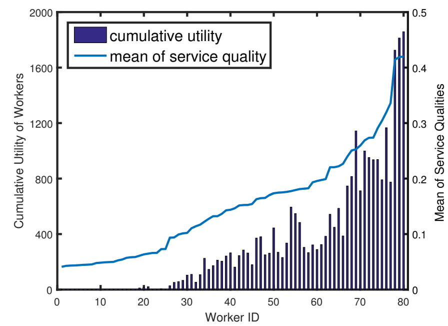

Fig. 6: Requester's cumulative utility with different context on MIT Reality.

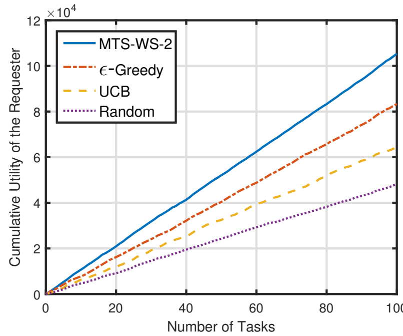

Fig. 7: Cumulative regret with different context on MIT Reality.

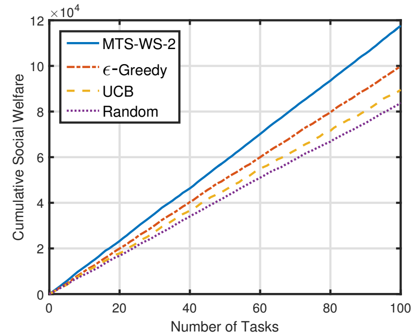

Fig. 8: Comparison of requester's cumulative utility on MIT Reality.

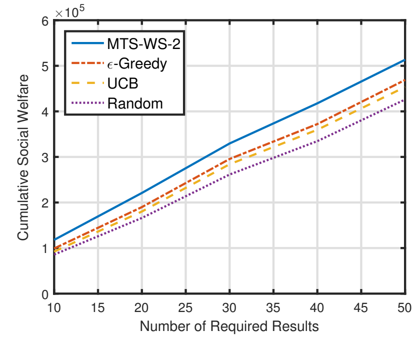

Fig. 9: Comparison of cumulative social welfare on MIT Reality.

tower IDs of w_i during a selection round as its activity scope.
- Encounter level: We use the number of different device IDs discovered by the Bluetooth of w_i during a selection round as its encounter level.

We evaluate the performance of our MTS-WS algorithm when different dynamic context information of a worker is available. Table II shows the names of the MTS-WS algorithm under specific context.

Figs. 4 and 5 show the influence of different dynamic context on our MTS-WS algorithm. With different features, MTS-WS chooses different groups of workers to execute sensing tasks. We can see that MTS-WS-2 achieves the highest cumulative utility of the requester and the lowest cumulative regret. Hence, we choose the dynamic context used in MTS-WS-2 for the following evaluations. In addition, Fig. 5 shows the growth rates of regret become flat eventually, which verifies that our MTS-WS is bounded.

Fig. 6 depicts the relationship between cumulative utility and the mean of service quality of the 80 workers under the MTS-WS-2 algorithm. In this figure, workers are sorted by their mean of service qualities in an increasing order. It is clear that workers with relatively high service quality can achieve considerable utilities. And the workers with relatively low service quality can hardly get any benefits. This phenomenon exemplifies that workers with high service quality are more frequently selected to execute tasks, while low quality ones are not.

Figs. 7 and 8 show the cumulative utility of the requester and the cumulative social welfare. The conclusion is similar to the results in the simulated data scenarios. Our MTS-WS achieves the highest cumulative utility of the requester and the social warfare.

Fig. 9 shows the comparison of the cumulative social welfare under different numbers of required results. In this figure, workers are chosen to execute 100 sensing tasks. We can see that our MTS-WS achieves the highest cumulative social warfare no matter how many results are needed.

In summary, the simulation results show that: i) MTS-WS algorithm can more accurately learn a worker's service quality by using the worker's historical and current context information; ii) When the context information changes, MTS-WS can dynamically adapt to select the suitable workers; and, iii) MTS-WS achieves the highest requester utility and the social welfare compared to the other three algorithms.

VII. Conclusion

In this paper, we proposed a Context-aware Multi-Armed Bandit incentive mechanism to facilitate quality-based worker selection in an MCS system. We extended the context information of a worker to include both extrinsic and intrinsic abilities so to more accurately evaluate the worker's quality. Based our proposed C-MAB and quality evaluation parameters, we developed a modified Thompson Sampling worker selection algorithm to select workers in a reinforcement learning manner. We theoretically proved that our C-MAB incentive mechanism was selection efficient, computationally efficient, individually rational, and truthful. Finally, we evaluated our MTS-WS algorithm in comparison with some other classic algorithms. Our evaluation results demonstrated that MTS-WS achieved the highest cumulative utility of the requester and social welfare. The proposed learning based MTS-WS can further enable efficient MCS systems for a wide range of smart city applications.

Acknowledgment

The work of Fan Li is partially supported by the NSFC (No. 61772077, 61370192), and the Beijing Natural Science Foundation (No. 4192051). The work of Liran Ma is partially supported by the US National Science Foundation (No. OAC-1829553). The work of Yu Wang is partially supported by the NSFC (No. 61572347), the US National Science Foundation (No. CNS-1343355), and by the U.S. Department of Transportation Center for Advanced Multimodal Mobility Solutions and Education (No. 69A3351747133). F. Li is the corresponding author.

References

[1] G. Li, J. Wang, Y. Zheng, and M. J. Franklin, "Crowdsourced data management: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 9, pp. 2296–2319, 2016.

[2] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, "Incentives for mobile crowd sensing: A survey," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 1, pp. 54–67, 2016.

[3] V. Coric and M. Gruteser, "Crowdsensing maps of on-street parking spaces," in *DCOSS*, 2013, pp. 115–122.

[4] X. Wang, J. Zhang, X. Tian, X. Gan, Y. Guan, and X. Wang, "Crowdsensing-based consensus incident report for road traffic acquisition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 8, pp. 2536–2547, 2018.

[5] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrasekaran, W. Xue, M. Gruteser, and W. Trappe, "ParkNet: drive-by sensing of road-side parking statistics," in *ACM MobiSys*, 2010, pp. 123–136.

[6] Y. Liu and M. Liu, "An online learning approach to improving the quality of crowd-sourcing," in *ACM SIGMETRICS*, 2015, pp. 217–230.

[7] P. Rudiger, R. Manfred, L. Berthold, and S. Winfried, "Mobile crowd sensing services for tinnitus assessment, therapy, and research," in *IEEE MS*, 2015.

[8] T. Li, T. Jung, H. Li, L. Cao, W. Wang, X. Li, and Y. Wang, "Scalable privacy-preserving participant selection in mobile crowd sensing," in *IEEE PerCom*, 2017, pp. 59–68.

[9] H. To, G. Ghinita, and C. Shahabi, "A framework for protecting worker location privacy in spatial crowdsourcing," *very large data bases*, vol. 7, no. 10, pp. 919–930, 2014.

[10] H. Xiong, D. Zhang, G. Chen, L. Wang, and V. Gauthier, "CrowdTasker: Maximizing coverage quality in piggyback crowdsensing under budget constraint," *ACM Transactions on Intelligent Systems and Technology*, vol. 6, no. 3, pp. 39–39, 2015.

[11] D. Zhang, H. Xiong, L. Wang, and G. Chen, "CrowdRecruiter: selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *ACM UbiComp*, 2014, pp. 703–714.

[12] H. Li, T. Li, and Y. Wang, "Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks," in *MASS*, 2015, pp. 136–144.

[13] X. Wang, Z. Liu, X. Tian, X. Gan, Y. Guan, and X. Wang, "Incentivizing crowdsensing with location-privacy preserving," *IEEE Transactions on Wireless Communications*, vol. 16, no. 10, pp. 6940–6952, 2017.

[14] H. Li, T. Li, F. Li, S. Yang, and Y. Wang, "Multi-expertise aware participant selection in mobile crowd sensing via online learning," in *IEEE MASS*, 2018.

[15] Y. Wang, D. Xu, and F. Li, "Providing location-aware location privacy protection for mobile location-based services," *Tsinghua Science and Technology*, vol. 21, no. 3, pp. 243–259, 2016.

[16] L. Cheng, C. Chen, J. Ma, and Y. Chen, "A group-level incentive scheme for data collection in wireless sensor networks," in *IEEE CCNC*, 2009, pp. 1–5.

[17] H. Li, T. Li, F. Li, W. Wang, and Y. Wang, "Enhancing participant selection through caching in mobile crowd sensing," in *IEEE IWQoS*, 2016, pp. 1–10.

[18] M. N. Katehakis and A. F. Veinott, "The multi-armed bandit problem: decomposition and computation," *Mathematics of Operations Research*, vol. 12, no. 2, pp. 262–268, 1987.

[19] S. Agrawal and N. Goyal, "Analysis of thompson sampling for the multi-armed bandit problem," in *ICLT*, 2012, pp. 39–1.

[20] A. Farshad, M. K. Marina, and F. Garcia, "Urban WiFi characterization via mobile crowdsensing," in *IEEE/IFIP NOMS*, 2014, pp. 1–9.

[21] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, 2011.

[22] G. Cardone, L. Foschini, P. Bellavista, A. Corradi, C. Borcea, M. Talasila, and R. Curtmola, "Fostering participaction in smart cities: a geo-social crowdsensing platform," *IEEE Communications Magazine*, vol. 51, no. 6, pp. 112–119, 2013.

[23] 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, pp. 1231–1239.

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

[25] Y. Liu, B. Guo, Y. Wang, W. Wu, Z. Yu, and D. Zhang, "TaskMe: multi-task allocation in mobile crowd sensing," in *ACM UbiComp*, 2016, pp. 403–414.

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

[27] Y. Wen, J. Shi, Q. Zhang, X. Tian, Z. Huang, H. Yu, Y. Cheng, and X. Shen, "Quality-driven auction-based incentive mechanism for mobile crowd sensing," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 9, pp. 4203–4214, 2015.

[28] S. Liu, Z. Zheng, F. Wu, S. Tang, and G. Chen, "Context-aware data quality estimation in mobile crowdsensing," in *IEEE INFOCOM*, 2017.

[29] B. Ye, Y. Wang, and L. Liu, "Crowd Trust: A context-aware trust model for worker selection in crowdsourcing environments," in *IEEE ICWS*, 2015, pp. 121–128.

[30] D. A. Berry and B. Fristedt, *Bandit problems: sequential allocation of experiments*. Springe Science, 2013.

[31] H. Robbins, "Some aspects of the sequential design of experiments," *Bulletin of the American Mathematical Society*, vol. 58, no. 5, pp. 527–535, 1952.

[32] W. R. Thompson, "On the likelihood that one unknown probability exceeds another in view of the evidence of two samples," *Biometrika*, vol. 25, no. 3/4, pp. 285–294, 1933.

[33] M. Tokic, "Adaptive ε-greedy exploration in reinforcement learning based on value differences," in *AAAI*, 2010, pp. 203–210.

[34] R. Agrawal, "Sample mean based index policies with $o(logn)$ regret for the multi-armed bandit problem," *Advances in Applied Probability*, vol. 27, no. 4, pp. 1054–1078, 1995.

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

[36] T. L. Lai and H. Robbins, "Asymptotically efficient adaptive allocation rules," *Advances in applied mathematics*, vol. 6, no. 1, pp. 4–22, 1985.

[37] P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, "The nonstochastic multiarmed bandit problem," *SIAM Journal on Computing*, vol. 32, no. 1, pp. 48–77, 2002.

[38] J. Langford and T. Zhang, "The epoch-greedy algorithm for multi-armed bandits with side information," in *NIPS*, 2008, pp. 817–824.

[39] Y. Gai, B. Krishnamachari, and R. Jain, "Combinatorial network optimization with unknown variables: multi-armed bandits with linear rewards and individual observations," *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1466–1478, 2012.

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

[41] S. Wang and W. Chen, "Thompson sampling for combinatorial semi-bandits," in *ICML*, 2018, pp. 5101–5109.

[42] U. U. Hassan and E. Curry, "A multi-armed bandit approach to online spatial task assignment," in *UIC*, 2014, pp. 212–219.

[43] L. Tranthanh, S. Stein, A. Rogers, and N. R. Jennings, “Efficient crowdsourcing of unknown experts using bounded multi-armed bandits,” *Artificial Intelligence*, vol. 214, no. 1, pp. 89–111, 2014.

[44] H. Zhang and M. Sugiyama, “Task selection for bandit-based task assignment in heterogeneous crowdsourcing,” in *TAAI*, 2015, pp. 164–171.

[45] N. Eagle, A. Pentland, and D. Lazer, “Inferring social network structure using mobile phone data,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 36, pp. 15 274–15 278, 2009.

Yue Wu received the BE degree in Internet of things from Beijing Institute of Technology, China in 2015. Currently she is a Ph.D. candidate in the School of Computer Science, Beijing Institute of Technology, Beijing, China. Her research interests include mobile crowd sensing, edge computing, and deep learning.

Fan Li is a Professor at the School of Computer Science in Beijing Institute of Technology, China. She received her PhD degree in Computer Science from the University of North Carolina at Charlotte, MEng degree in Electrical Engineering from the University of Delaware, MEng and BEng degrees in communications and information system from Huazhong University of Science and Technology, China, respectively. Her current research focuses on wireless networks, smart sensing, crowd sensing and mobile computing. Her papers won Best Paper Awards from IEEE MASS (2013), IEEE IPCCC (2013), ACM MobiHoc (2014), and Tsinghua Science and Technology (2015). She is a member of ACM and IEEE.

Liran Ma is currently an Associate Professor in the Department of Computer Science at Texas Christian University. He received his D.Sc. degree in Computer Science from the George Washington University. His current research focuses on wireless, mobile, and embedded systems, including security and privacy, smartphones, smart health, mobile computing, data analytics, security and privacy, Internet of Things, and cloud computing. It involves building and simulating prototype systems and conducting real experiments and measurements.

Yadong Xie received the BE degree in network engineering from Hebei University, China in 2016. Currently he is a Ph.D. candidate in the School of Computer Science, Beijing Institute of Technology, Beijing, China. His research interests include Internet of vehicles, mobile computing, and machine learning.

Ting Li received the B.S. degree in information engineering from Xidian University, China, in 2012. She is currently a Ph.D. candidate in computer science at the University of North Carolina at Charlotte. Her research interests include security and privacy, mobile crowd sensing, and wireless.

Yu Wang is a Professor of Computer Science at the University of North Carolina at Charlotte. He received his Ph.D. degree in Computer Science from Illinois Institute of Technology, his B.Eng. degree and M.Eng. degree in Computer Science from Tsinghua University, China. His research interest includes wireless networks, mobile social networks, smart sensing, mobile crowd sensing, mobile computing, and algorithm design. His research has been continuously supported by federal agencies including US National Science Foundation, US Department of Transportation, and National Natural Science Foundation of China (NSFC). He has published over 150 papers in peer reviewed journals and conferences, with four best paper awards. He is a recipient of Ralph E. Powe Junior Faculty Enhancement Awards from Oak Ridge Associated Universities (2006), Outstanding Faculty Research Award from College of Computing and Informatics at UNC Charlotte (2008), and Overseas Young Scholars Cooperation Research Fund from NSFC (2014). He is a senior member of the ACM and a fellow of the IEEE.