Energy-Aware Participant Selection for Smartphone-Enabled Mobile Crowd Sensing

Chi Harold Liu, *Member, IEEE*, Bo Zhang, Xin Su, *Member, IEEE*, Jian Ma, Wendong Wang, *Member, IEEE*, and Kin K. Leung, *Fellow, IEEE*

***Abstract*—Mobile crowd sensing systems have been widely used in various domains but are currently facing new challenges. On one hand, the increasingly complex services need a large number of participants to satisfy their demand for sensory data with multidimensional high quality-of-information (QoI) requirements. On the other hand, the willingness of their participation is not always at a high level due to the energy consumption and its impacts on their regular activities. In this paper, we introduce a new metric, called "QoI satisfaction ratio," to quantify how much collected sensory data can satisfy a multidimensional task's QoI requirements in terms of data granularity and quantity. Furthermore, we propose a participant sampling behavior model to quantify the relationship between the initial energy and the participation of participants. Finally, we present a QoI-aware energy-efficient participant selection approach to provide a suboptimal solution to the defined optimization problem. Finally, we have compared our proposed scheme with existing methods via extensive simulations based on the real movement traces of ordinary citizens in Beijing. Extensive simulation results well justify the effectiveness and robustness of our approach.**

***Index Terms*—Energy efficiency, mobile crowd sensing (MCS), participant selection, sampling behavior.**

I. Introduction

With the development of mobile Internet, mobile crowd sensing (MCS) has become popular in recent years [1]. The key idea of MCS is to recruit ordinary citizens to collect and share sensory data from their surrounding environment by using their energy-constrained smart devices [2], such as iPad, smartphone, and Google Glass. Different from other sensing systems (i.e., participatory sensing), mobile social networks and mobile sensing are widely used in MCS.

Our research in this paper is motivated by an application scenario as shown in Fig. 1, where a group of mobile users subscribes to a central server, who receives sensing tasks from either local or remote, online task publishers. A task is associated with multidimensional quality-of-information (QoI) requirements, including, but not limited to, the sensing region, time period, data granularity quantity, and the affordable incentive budget. Without loss of generality, we assume that a sensing region is a 2-D area and is divided into multiple virtual subareas of the same size based on the spatiotemporal granularity requirement of a task. A task lasts for a certain period of time, which is also divided into discrete time slots accordingly. In each time slot on each area, a certain number of measurements (i.e., sensory data samples) are requested by tasks. On the participant's side, they move around in the sensing region, and when they receive a task and its associated QoI requirements, they contact the central server to report their current GPS location, required amount of incentives, remaining energy level of their device, and their sensing capabilities on the carried smart device. In our considered scenario, the sensing capability of a participant is measured by how many data samples he/she can collect in a time unit, determined by both the sampling interval and the type/amount of sensors equipped on his/her devices. To minimize the total energy consumption of all participants, we assume that they can take samples with different sampling intervals; for example, the participant carrying a low energy level device can reduce his/her sampling interval to save power.

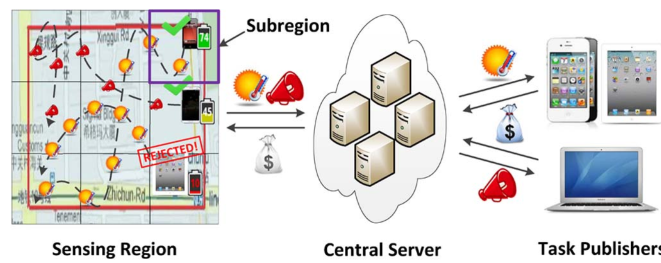

Fig. 1. Considered MCS scenario to monitor the fire and extreme temperatures in a city. A subset of all candidate participants is selected from all users with different remaining energy levels and initial locations.

When a particular participant is selected by the central server as a data collector, his/her smart device measures the requested environmental parameters periodically and uploads the collected data to the server. This data collection phase stops either when all QoI requirements are fully satisfied, or no more participants can provide better sensory data, or when the given task budget runs out.

The participant selection scheme is critical to the efficient operation of the aforementioned MCS systems, since the high density of smart device users in urban areas makes it possible

Manuscript received January 15, 2015; revised April 9, 2015; accepted April 25, 2015. This work was supported by the National Natural Science Foundation of China under Grant 61300179. *(Corresponding author: Chi Harold Liu.)*

C. H. Liu is with the School of Software, Beijing Institute of Technology, Beijing 100081, China (e-mail: chiliu@bit.edu.cn).

B. Zhang, J. Ma, and W. Wang are with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China (e-mail: isler@bupt.edu.cn; wdwang@bupt.edu.cn; majian@mwsn.com.cn).

X. Su is with the Research Institute of Information Technology, Tsinghua University, Beijing 100086, China (e-mail: suxin@tsinghua.edu.cn).

K. K. Leung is with the Department of Electrical and Electronic Engineering and the Department of Computing, Imperial College London, London SW7 2BT, U.K. (e-mail: kin.leung@imperial.ac.uk).

Digital Object Identifier 10.1109/JSYST.2015.2430362

to select only a subset of all available participants; however, different participant selection schemes may lead to different system performances. According to [3], the accuracy of sensing results is highly related to the number of collected samples, and thus, the server tends to recruit as many participants as possible to improve the QoI. On the other hand, collecting data from colocated users may result in redundant data that cannot help further improve QoI but may result in high usage of energy, network bandwidth, and storage space. Furthermore, tasks often have *time-varying* QoI requirements, e.g., sensing region and data quality vary from time to time. Then, the system cannot persistently recruit the same group participants covering the entire region. Toward this end, in this paper, we propose a novel MCS model, such that the system could meet the QoI requirements of tasks while all the participants feel no antipathy toward the sensing tasks caused by possible energy consumption.

Although several participant selection strategies have been proposed, they either assumed the participants' trajectories are known *a priori* and its fitness to the task's sensing region [4]–[13] or aimed to improve the energy efficiency by using collaborative sensing, or replacing high energy-cost sensors with low energy-cost sensors in specific application scenarios [14]–[17]. Different from all these, our goal is to select a crowd of participants to provide high-QoI satisfaction for multiple sensing tasks simultaneously, while minimizing the overall energy consumption. The contribution of this paper is fourfold.

1) We introduce a novel concept of "QoI satisfaction ratio" to quantify the degree of how collected sensory data can satisfy multidimensional QoI requirements of tasks in terms of data granularity and quantity.
2) We propose a novel participant sampling behavior model, to quantify and explicitly build up the relationship between their remaining energy level and the willingness for participation, given that helping out a sensing task may impact their ordinary device usage. Based on this, we calculate the rejection probability that represents the chance of a participant to reject the sensing task if the recommended number of data samples from the server cloud exceeds his/her sensing capabilities.
3) We formulate a constrained optimization problem to select participants in an energy-efficient and QoI-aware manner. Then, we propose a suboptimal solution to solve this problem.
4) The effectiveness and flexibility of the proposed strategy have been extensively evaluated by real trace-driven simulations.

The rest of this paper is organized as follows. Section II reviews the related research activities. Section III establishes formal models of our systems. Section IV describes the proposed QoI satisfaction ratio and builds up its connection with energy consumption and willingness for participation. Section V introduces the definition of our considered optimization problem and provides a suboptimal solution. Section VI extensively evaluates the performance of the proposed strategy by real trace-driven simulations, and finally, Section VII concludes this paper.

II. Related Work

The concept of participatory sensing was first proposed in [18] and then followed by many proposals. For example, the Common Sense project [19] develops a participatory sensing system that allows individuals to measure their personal exposure to air pollution, and Rana *et al.* in [20] presented a system that collects and shares the noise pollution information. Kanjo in [21] first studied the platforms for multiple sensing tasks and proposed a procedural programming language for collecting multiple types of sensor data from a large number of mobile phones. MEDUSA [22] synthesizes participatory sensing and crowdsourcing and puts forward a runtime system for multiple sensing tasks with the following stages: task submission, worker selection, and monetary incentive management. In [4], Reddy *et al.* developed a selection framework to enable organizers to identify well-suited participants for data collection, based on both geographic and temporal availability as well as participation habits. In [5], Tuncay *et al.* exploited the stability of user behaviors and proposed to select participants based on the fitness of mobility history profiles. Similarly, Weinschrott *et al.* in [6] and Zhong and Cassandras in [7] discussed the task assignment problem for opportunistic *in situ* sensing, and the research in [9] focuses on initiating sampling around specific location "bubbles" (i.e., regions). Gaonkar *et al.* in [10] proposed a coverage maximization algorithm that records participants' tracks and selects participants whose availability matches the campaign coverage constraints. These schemes highly rely on the knowledge of participant trajectories and, thus, may lead to the increased risk of mobile users' privacy leakage. Riahi *et al.* in [11] and Duan in [12] further took into account the incentive request of participants, and they aimed mainly at how to optimize data acquisition quality by allocating limited amount of incentive.

Many researchers have studied the issue of trajectory prediction as the location-acquisition technologies (GPS, GSM networks, etc.) are getting mature. The conclusion that a human behaves a certain degree of temporal and spatial regularity following simple and reproducible patterns proposed in [23] and [24] shows a considerable predictable degree of human behavior. Moreover, it makes those trajectory prediction methods reasonable and feasible. Zhang *et al.* in [25] investigated the large-scale user mobility traces that are collected by a telecom operator. Gambs *et al.* in [26] proposed a mobility model called mobility Markov chain (MMC), which incorporates the n previous visited locations. In [27] and [28], Ruan *et al.* and Wu *et al.*, respectively, used the Markov model to predict moving trajectories of participants.

Zhang *et al.* in [29] attempted to fill the gap by dividing the life cycle of the MCS process into four stages and using "4W1H" (i.e., what/when/where/who/how) to characterize the major research issues in each of the four stages of the MCS life cycle as well as across the whole MCS process. Zhang *et al.* in [30] designed a novel community-centric framework for community activity prediction based on big data analysis and proposed an approach to extract community activity patterns by analyzing the big data collected from both the physical world and virtual social space. Cardone *et al.* in [31] initialized a project, called "ParticipAct Living Lab testbed," as an ongoing experiment at the University of Bologna involving 300 students for one year in crowd sensing campaigns that can passively access smartphone sensors and also require active user collaboration. Pankratius *et al.* in [32] discussed an application of crowdsourcing in space weather monitoring, called "the Mahali project." Mahali used GPS signals that penetrate the ionosphere for science rather than positioning. A large number of ground-based sensors will be able to feed data through mobile devices into a cloud-based processing environment, enabling a tomographic analysis of the global ionosphere at unprecedented resolution and coverage.

Participant selection in multitask systems is quite different than that in single-task systems. In [12], assuming that incentive requests of participants and the utility of sensory data on all locations are known, Duan proposed to select a subset of participants with maximum sensory data utility deducting incentive requirements. In [11], Riahi *et al.* further improved [12] by studying how to calculate sensory data utility on a certain location. Both research works concentrate on selecting participants to maximize the difference between value and price of sensory data.

Finally, energy efficiency has recently been investigated for MCS. Baier *et al.* in [15] and Nath in [16] improved device battery lifetimes by inferring sensor readings by sensors with lower energy consumption or temporal continuous readings, since the values of various context attributes can be highly correlated. Sheng *et al.* in [17] proposed several minimum energy sensing scheduling algorithms. Different from all these, our proposed participant selection scheme in this paper can efficiently and dynamically select a desirable set of participants to meet the required QoI levels with a limited task budget while fully considering the impact of energy consumption on their willingness for participation.

III. System Model

This section presents a formal model for our system. We consider a multitask-oriented MCS system in a 2-D region $\mathcal{L}$, as shown in Fig. 1. The system is composed of a set of task publishers, a central server, a set of M smart device carriers $\mathcal{M} \triangleq \{m = 1, 2, \ldots, M\}$, and all of them are associated with a set of sensing tasks $\mathcal{Q} \triangleq \{q = 1, 2, \ldots, Q\}$.

When a sensing task q and its budget c_q is received, the central server gathers the whole budget from all tasks, which is denoted by C, then the central server selects a set of participants as the sensory data contributors from candidate participants, who are moving in the region $\mathcal{L}$. The region $\mathcal{L}$ is divided into a set of l subregions, denoted by $\mathcal{L} \triangleq \{l = 1, 2, \ldots, L\}$. The region inside the boundary of each task q is denoted by $\mathcal{L}_q, \forall q \in \mathcal{Q}, \mathcal{L}_q \subseteq \mathcal{L}$. The sensing task will last for t time slots, denoted by $\mathcal{T} = \{t = 1, 2, 3, \ldots, T\}$. Then, the participants need to upload the sensory readings from their devices and finally received their incentive as a reward for their participation. In each time slot and in each subregion, a minimum of r_{lt}^q samplings are required to obtain accurate sensing results. Without loss of generality, we also assume that sensing tasks are independent of each other, both temporarily and spatially, as shown in Fig. 2.

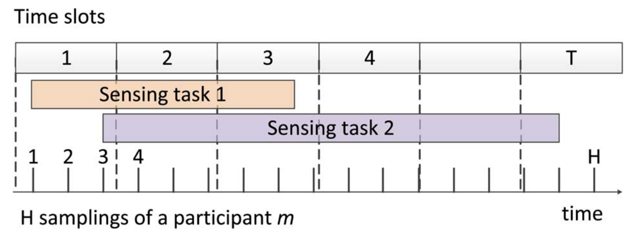

Fig. 2. Time distribution of different sensing tasks.

During $\mathcal{T}$, when a participant m is selected as a data contributor, his/her device then samples the required environmental parameters, such as fire and temperature, periodically by the equipped sensor(s). Then, $s_m^q, \forall m \in \mathcal{M}, \forall q \in \mathcal{Q}$ is used to denote the sensing capability of participant m to task q, where

$$s_m^q = \begin{cases} 0, & \text{if } m \text{ cannot collect data for } q \\ 1, & \text{otherwise.} \end{cases} \tag{1}$$

In MCS, data samples (i.e., sensor readings), such as temperature, humidity, noise levels, etc., could be taken within a few milliseconds, which is much less than the duration of a time slot, so that we can safely assume that the effect of sampling time could be ignored. We also assume that each candidate participant's smart device has an initial remaining energy when entering the sensing region, denoted by $e_m, \forall m \in \mathcal{M}$.

Since the trajectories of participants are unknown as *a priori* when they enter the sensing region, the system needs to estimate the potential data contributions from these participants during their future movement, so that their potential data contributions can be computed to aid the participant selection decision. We assume that only the initial locations of each participant $m \in \mathcal{M}$ is known when entering the sensory region, denoted by $E_m(0)$, which can be inferred from their uploaded GPS records; however, their future trajectories are *not* known but can be estimated. The general procedure of trajectory prediction consists of two stages:

1) **Step 1.** Training the prediction model based on historical trajectory data.
2) **Step 2.** Taking the trained model and the current movements of individuals needed to be predicted as input to predict the future movements.

Here, we use a kth-order Markov chain as the trajectory prediction model, which is introduced in [33]. Let $P(n)$ denote the n-step transition probability matrix for a particular participant and $p_{l_1 l_2}$ denote the probability of that participant moving from area l_i to l_j, $\forall i, j \in L$. Then, given his/her historical trajectory information, we are able to calculate the probability of a participant moves from l_i to l_j, denoted by N_{ij}, and $p_{l_1 l_2}$ can be calculated as follows:

$$p_{l_1 l_2} = \frac{N_{ij}}{\sum_{j=1}^{n} N_{ij}} \tag{2}$$

and the one-step transition probability matrix can be denoted by

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1L} \\ p_{21} & p_{22} & & \\ \cdots & & \cdots & \\ p_{L1} & & & p_{LL} \end{bmatrix}. \tag{3}$$

TABLE I
LIST OF IMPORTANT NOTATIONS AND THEIR DESCRIPTIONS

Notation	Explanation
$\mathcal{Q}$	A set of sensing tasks being serviced, where each task is q
$\mathcal{M}$	A set of participants
$\mathcal{L}_q$	Sensing area division by task q
$\mathcal{T}_q$	Time slot division by task q
$\underline{R}^q$	Data requirement of task q, where r_{lt}^q denotes detailed requirement on area l and time t
$\underline{O}^q(\mathcal{X})$	Collected data by $\mathcal{X}$ for task q
$\underline{U}(\mathcal{X})$	QoI satisfaction vector achieved by $\mathcal{X}$, where task q's index is $u^q(\mathcal{X})$
g^q	The QoI requirement given by task q
e_m	Initial energy of participant m
C	Amount of budget given by all tasks, where task q's budget is c_q
d_m	Amount of incentive required by participant m
b_m	The recommend number of data samples to participant m
P	Position transition matrix
$\underline{E}_m(t)$	Position matrix of participant m at time t, where his initial location is $\underline{E}_m(0)$

Then, we can calculate the n-step transition probability matrix, dented as $P(n) = P^n$. In our model, we use the last k positions and $P(k)$ to estimate the future locations of a participant, which is described as the prediction step in the following sections. Since the most recent position could have a higher impact than previous positions when estimating their future trajectory, we simply weight the historical data as

$$x(i+1) = \sum_{j=i-k}^{i} a_j x(i-j+1)P^j \tag{4}$$

where $x(i+1)$ denotes the probability matrix of the next position, a_k denotes the weight, and we set $a_1 > a_2 > \cdots > a_k$. Next, we can calculate the probabilities of a participant moving from the current position to all other positions in the discretized sensing region, and we simply use the position with highest probability as the predicted point that he/she will most likely visit in the next time slot. Iteratively, the entire future trajectory can be calculated for all participants in the sensing region.

Table I shows the list of important notations used in this paper.

IV. Energy-Aware QoI Model

Here, we first introduce a novel concept of "QoI satisfaction ratio" as an index to measure to what degree the QoI requirements of a task is satisfied. Then, we propose a novel scheme to map the remaining energy level to recommended sampling behavior and study the potential task rejection probability due to limited device resource.

A. QoI Satisfaction Ratio

As its name implies, the QoI satisfaction ratio is used to describe the level of QoI satisfaction that the collected sensory data can provide to the QoI requirements of a task. When a sensing task is published, there should be some limit for the required sensory data. Generally, the limit consists of two parts: the total number and the distribution requirement of the sensory data. Thus, when we formulate the QoI satisfaction ratio, we should consider these two aspects. On one hand, based on MCS, the total number of the sensory data could be satisfied by the large number of participants. On the other hand, with the division of the sensing region, if we can collect "enough" sensory data in each subregion, we can consider that the distribution requirement is satisfied.

In our considered scenario, a subset $\mathcal{X}$ (of size $|\mathcal{X}|$) of all participants is selected for task q, and let $o_{lt}^q(\mathcal{X})$ denote the number of data samples collected by a group of participant $\mathcal{X}$ for task q on a certain area l, at time slot t. The initial value of each $o_{lt}^q(\mathcal{X})$ is set to zero. When a new data sample on area l at time slot t is collected, if the amount of collected data $o_{lt}^q(\mathcal{X})$ is less than the amount of required data r_{lt}^q, then $o_{lt}^q(\mathcal{X})$ is increased by 1; otherwise, if the amount of collected data $o_{lt}^q(\mathcal{X})$ has reached the required amount of data, then r_{lt}^q, $o_{lt}^q(\mathcal{X})$ do not change.

Therefore, we use two matrices, $\underline{R}^q$ and $\underline{O}(\mathcal{X})^q$, to denote the QoI requirements of task q, and the number of data sample collected by $\mathcal{X}$, respectively, as

$$\underline{R}^q = \begin{bmatrix} r_{11}^q & r_{12}^q & \cdots & r_{1T}^q \\ r_{21}^q & r_{22}^q & & \\ \cdots & & \cdots & \\ r_{L1}^q & & & r_{LT}^q \end{bmatrix} \tag{5}$$

$$\underline{O}^q(\mathcal{X}) = \begin{bmatrix} o_{11}^q(\mathcal{X}) & o_{12}^q(\mathcal{X}) & \cdots & o_{1T}^q(\mathcal{X}) \\ o_{21}^q(\mathcal{X}) & o_{22}^q(\mathcal{X}) & & \\ \cdots & & \cdots & \\ o_{L1}^q(\mathcal{X}) & & & o_{LT}^q(\mathcal{X}) \end{bmatrix} \tag{6}$$

where $\forall\, l \in \mathcal{L}, t \in \mathcal{T}$, we have

$$o_{lt}^q(\mathcal{X}) = o_{lt}^q\left(\sum_{\forall\, m\in\mathcal{X}} m\right). \tag{7}$$

To meet a task's multidimensional QoI requirements, an index, as the proposed QoI satisfaction ratio in this paper, is used to explicitly quantify the degree of QoI satisfactions for task q in area l and time slot t, as

$$u_{lt}^q = \min\left(1, \frac{o_{lt}^q}{r_{lt}^q}\right) \in [0,1], \qquad \forall\, q \in \mathcal{Q}. \tag{8}$$

Then, we use $u^q(\mathcal{X})$ to represent the total achieved level of QoI satisfactions for task q by a group of participants $\mathcal{X}$ as

$$u^q(\mathcal{X}) = \frac{\sum_{\forall\, l\in\mathcal{L}_q, \forall\, t\in\mathcal{T}_q} u_{lt}^q}{\mathcal{L}_q \cdot \mathcal{T}_q}, \qquad \forall\, q \in \mathcal{Q}, \mathcal{X} \subseteq \mathcal{M} \tag{9}$$

and we call $u^q(\mathcal{X})$ as the "QoI satisfaction ratio" of task q; it ranges from 0 to 1. While lower bound 0 indicates that no data have been collected for task q, upper bound value 1 means that all QoI requirements at each area and within each time slot are fully satisfied. If the collected data $\underline{O}^q(\mathcal{X})$ do not meet the QoI requirement matrix $\underline{R}^q$, this QoI satisfaction ratio may further increase when more data are collected. Then, task publishers can request a certain level of QoI satisfaction ratio, when publishing sensing asks, denoted by $g^q, \forall\, q \in \mathcal{Q}$, and the default value is 1 (fully satisfied).

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

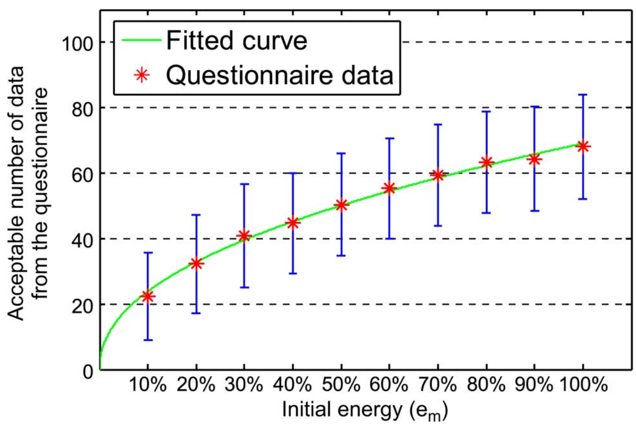

Fig. 3. Fitting result of α, β by using the data obtained from an online questionnaire.

B. Energy-Aware Recommended Sampling Behavior

We consider the impact of data collection on the participants' regular activities, i.e., the energy consumption and the disturbance by frequent samplings. Here, the goal is to study the relationship between initial energy and recommended sampling behavior and ultimately calculate the task rejection probability by participants due to their limited device energy.

We denote the recommended number of data samples for a participant as $b_m, \forall m \in \mathcal{M}, \forall q \in \mathcal{Q}$, based on his/her remaining energy level and initial location. A lower b_m value indicates that if the smart device's energy of the participant is in a low level, the central server will suggest him/her to collect less sensory data, so that the participant may be more likely to accept the sensing task. On the contrary, if the smart device's energy of the participant is in a high level, the central server will recommend a higher b_m. Thus, if every participant can receive an *appropriate* recommended number of data samples, he/she would be more likely to accept the sensing task, and the central server also can obtain satisfactory sensory data. Thus, b_m should be set to a higher value when the initial energy e_m is in a higher level. Then, the relationship between the *appropriate* b_m and the initial energy e_m of participant m can be established by a generic formulation $b_m = y(e_m)$, where $y : \mathbb{R} \to \mathbb{R}$.

To obtain the mapping function $y(\cdot)$, we conduct an online questionnaire. We specified a few questions including: "What is the maximum suitable number of data samples that you are willing to contribute, to the environmental sensory data gathering, when your remaining energy is $(10\%, 20\%, \ldots, 100\%)$." Then, volunteers tick the choices $(0, 10, 20, \ldots, 100)$ data samples of their wish. Because the central server could not get the property of each participant in the sensing task, to reflect the participants' property of randomization, we do not limit the scope of the volunteers. However, we limit that the online questionnaire can be answered once for each IP, so that we will not receive too much repetitive survey results. As a result, 130 questionnaires are returned. The amount of volunteers is not many; thus, some online tools such as Amazon's Mechanical Turk can be used for more survey results. As shown in Fig. 3, we use the mean square error to demonstrate the upper and lower bounds of the collected statistics, and also the averaged value. By observing the curve connected by average values of the statistics at points $(10\%, 20\%, \ldots, 100\%)$, we found that we can use the following formulation for y as

$$b_m \triangleq \alpha(e_m)^{\beta}, \qquad \forall m \in \mathcal{M}. \tag{10}$$

By fitting parameters α, β, we obtain that $\alpha = 8.179$, $\beta = 0.4633$, and the fitting coefficient is 0.9951. We use them for the following derivation and experiments. Nevertheless, it is worth noting that for different MCS scenarios, different α, β values can be chosen by conducting similar empirical studies.

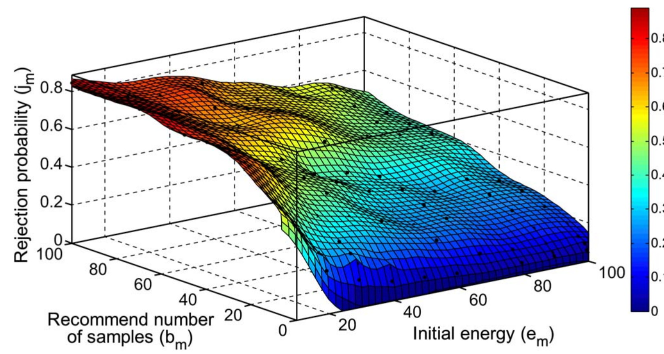

Fig. 4. Fitting result of j_m, b_m, e_m by using the data obtained from the online questionnaire.

The average values of the statistics seem to be a solution for the *appropriate* recommended number of data samples. However, based on the curve connected by average values of the statistics at points $(10\%, 20\%, \ldots, 100\%)$, we observe that if the central server requires the same recommended number of data samples for the participants with the same remaining energy level of their devices, some participants may reject the sensing task due to their devices' limited remaining energy level; however, they may appear at some less-visited locations of the sensing region, where other participants seldom visit. Therefore, allowing these participants to join the sensing tasks (rather than immediate rejection) is sometimes essential to improve the overall data quality. On the other hand, when too many participants appear in the same subregions, the central server could improve the value of b_m, so that the participants who accept the sensing task will upload more sensory data without any more incentive. In this case, the central server can save costs but will also face the risk of user rejection at the same time. To this end, we denote this "rejection probability" of a participant as $j_m, \forall m \in \mathcal{M}$, as a function of the remaining energy level of the participants' devices and the recommended number of data samples. Given the online questionnaire data, we are able to plot this mapping function $j_m = f(b_m, e_m)$, where $f : \mathbb{R}^2 \to \mathbb{R}$, as shown in Fig. 4. The relationship between j_m, b_m, and e_m can be established by $j_m = f(b_m, e_m)$, where $f : \mathbb{R}^2 \to \mathbb{R}$, and we found that we can use a cubic function to approximate it. By fitting parameters, we obtain a fitting coefficient of 0.9837. The curved surface denotes the value of rejection probability when the recommended number of data samples and the remaining energy level are certain. Each black point is an actual value calculated by the number of volunteers who choose a lower value than b_m when e_m is given and the total number of all volunteers. It is worth noting that the value of j_m will increase with increasing b_m or decreasing e_m.

V. QoI-Aware and Energy-Efficient Participant Selection

Here, we first formally define our considered constrained optimization problem and then analyze this problem and propose our suboptimal solution with low computational complexity.

A. Optimization Problem Definition

Our objective is to find an optimal crowd of participants to collect sufficient amount of sensory data with required data quality. Let $\mathcal{X}^*$ denote this optimal set of selected participants, and we formulate the following optimization problem:

$$\text{Maximize:} \quad u^q(\mathcal{X}) = \frac{\sum_{\forall l \in \mathcal{L}_q, t \in \mathcal{T}_q} u^q_{lt}}{\mathcal{L}_q \mathcal{T}_q}, \qquad \forall q \in \mathcal{Q}$$
$$\text{subject to:} \quad \sum_{\forall m \in \mathcal{X}} d_m \leq C, \qquad \mathcal{X} \subseteq \mathcal{M} \tag{11}$$

where $u^q(\mathcal{X})$ denotes the averaged achieved QoI satisfaction ratio on all areas of task's lifetime, d_m denotes the amount of requested incentives by participant m, and C denotes the incentive budget given by all tasks. Therefore, the goal of the above optimization problem is to maximize the achieved QoI satisfaction ratio of all tasks simultaneously, under the constraints of the limited incentive budget.

B. Proposed Suboptimal Solution

The objective function of (11) is similar to the knapsack problem (KP), where it maximizes a multiple-objective function subjected to one binary and linear capacity constraint. For a generic KP, there are a set of items, each with a weight, to determine the number of each item as to be included in a collection, so that the total weight is less than or equal to each given limit, and the total value is as large as possible. The general statement is

$$\text{Maximize: } f(x) \quad \text{subject to: } m(x) \leqslant p, \quad x \in \mathcal{S}. \tag{12}$$

Since KP is NP-complete, to obtain the optimal solution of (11), the number of iterations should be $\sum_{m=1}^{M} C_M^m$, so its complexity is $O(2^M)$, and thus, we are seeking an efficient suboptimal solution.

To solve a KP, the greedy algorithm is suitable; however, the standard way will not fit our scenario of multitask systems. To accelerate convergence, we propose a hybrid iterative greedy algorithm by weighting the tasks from different scales. Here, the weight for task q, denoted by ω^q, can be calculated as

$$\omega^q = \frac{1 - u^q}{\sum_{\forall q \in \mathcal{Q}} (1 - u^q)} \tag{13}$$

so that the QoI of all the tasks could be satisfied at the same time frame.

The steps of the proposed solution is as follows. First, we calculate the total contributions of each participant by summing up their weighted contributions of each task and then find the participant who is associated with the highest value. Second, we select participants one by one until the QoI requirements of all tasks are fully satisfied, or the task budget runs out. In such an iterative way, the proposed algorithm selects the most "efficient" participants. Here, we define the efficiency of a participant m in each round of iteration as

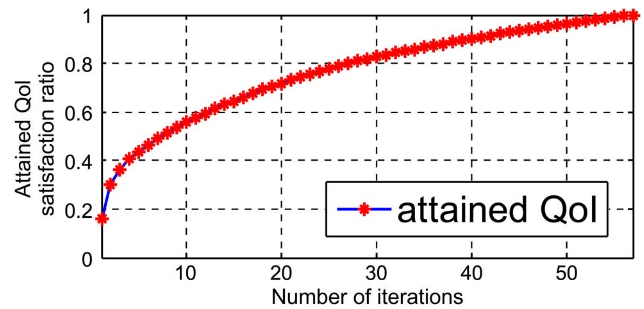

Fig. 5. Attained QoI satisfaction ratio over time.

$$\vartheta(m, \mathcal{X}') = \frac{1 - j_m}{d_m} \cdot \sum_{\forall q \in \mathcal{Q}} \omega^q \left(u^q(\mathcal{X}' + m) - u^q(\mathcal{X}')\right) \tag{14}$$

where $\mathcal{X}'$ denotes the set of participants that were selected in the previous round, and $\vartheta(m, \mathcal{X}')$ denotes the efficiency of a participant m in the current round. j_m is the rejection probability obtained in Section IV-B that discounts the calculated efficiency due to possible task rejections.

When the central sever requires the amount of collected sensory data b_m for participant m, where b_m is calculated based on (10), the central server will obtain the efficiency $\vartheta(m, \mathcal{X}')$ of a participant m in the current round. However, if the central server requires a different value of b_m, the value of $\vartheta(m, \mathcal{X}')$ should be changed. Obviously, with the growth of the recommended number of data samples, the amount of possible sensory data and the rejection probability of a participant will also increase; thus, there should be a highest value of $\vartheta(m, \mathcal{X}')$. Hence, we can calculate the maximum efficiency $\vartheta'(m, \mathcal{X}')$ of participant m in the current round as follows:

$$\vartheta'(m, \mathcal{X}') = \max\left(\vartheta(m, \mathcal{X}')\right), \qquad \forall b_m > 0. \tag{15}$$

Then, we use the number as the recommended number of data samples when we obtain $\vartheta'(m, \mathcal{X}')$.

Proposition 1: According to the definition of submodular functions [34], with more iteration steps, the attained QoI satisfaction ratio of a participant m will become attenuated, as: $\forall \mathcal{X}_1, \mathcal{X}_1 \subset \mathcal{M}, \forall m \in \mathcal{M}$, given $\mathcal{X}_1 \subset \mathcal{X}_2, m \notin \mathcal{X}_1, m \notin \mathcal{X}_2$, we have $\vartheta(m, \mathcal{X}'_2) \leq \vartheta(m, \mathcal{X}'_1)$.

Proposition 1 suggests that if we consider the participant selection scheme as a step-by-step procedure where they are iteratively selected, the achieved QoI satisfaction ratio is increased by recruiting a candidate m into a smaller group $\mathcal{X}_1$ than into a bigger group $\mathcal{X}_2$, if $|\mathcal{X}_1| < |\mathcal{X}_2|$. It therefore implies that the achieved QoI gain decreases by recruiting more participants, which will be verified by our observation in the experiment, as shown in Fig. 5.

Our proposed dynamic participant selection scheme runs at the beginning of each time period to select participants by a few

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

rounds of iterations. The pseudocode is given in Algorithm 1, and a detailed description is as follows.

Algorithm 1 Proposed solution for (11)

Require:
A set of sensing tasks $\mathcal{Q}$;
total task budget C; task budget left C_{left};
area and time division $\mathcal{L}^q, \mathcal{T}^q$;
required QoI satisfaction ratio g^q;
task QoI requirements $\underline{R}^q$;
A set of participants $\mathcal{M}$;
incentive requirement of each participant c_m;
sensing capability of each participant s_m^q;
initial locations of participants $\underline{E}_m(0), \forall m \in \mathcal{M}$;
interval between samples Δt;
transition matrix obtained from historic traces $\underline{P}$.

Ensure: Selected participants as set $\mathcal{X}^*$;
set of selected participants $\mathcal{A} = \text{NULL}$
set of unselected participants $\mathcal{B} = \mathcal{M}$
sort the unselected participants $\mathcal{B}$
while 1 **do**
  `flag` $\leftarrow 0$
  `qoiflag` $\leftarrow 0$
  `selected_id` $\leftarrow 0$
  `max_efficiency` $\leftarrow 0$
  for mobile user $m \in \mathcal{B}$ **do**
    compute m's max efficiency $\vartheta'(m, \mathcal{A})$ in (15)
    if $\vartheta'(m, \mathcal{A}) >$ `max_efficiency` **then**
      `selected_id` $\leftarrow m$
      `max_efficiency` $\leftarrow \vartheta(m)$
      `flag` $\leftarrow 1$
    end if
  end for
  if `flag` $= 0$ or `selected_id` $= 0$ or $c_m \geq C_{left}$ **then**
    break
  end if
  $\mathcal{A} \leftarrow \mathcal{A} +$ `selected_id`
  $\mathcal{B} \leftarrow \mathcal{B} -$ `selected_id`
  for task $q \in \mathcal{Q}$ **do**
    if $u^q(\mathcal{X}) < g^q$ **then**
      `qoiflag` $\leftarrow 1$
    end if
  end for
  if `qoiflag` $= 0$ **then**
    break
  end if
end while
Return: final selected participant set $\mathcal{X}^* = \mathcal{B}$.

1) **Step 1—Initialization**. At the beginning of the sensing time period, the participant selection strategy is initialized. All available participants are divided into two sets: the selected set $\mathcal{A}$ and the unselected set $\mathcal{B}$. In this step, all participants are put in $\mathcal{B}$, and $\mathcal{A}$ is set to $\emptyset$.

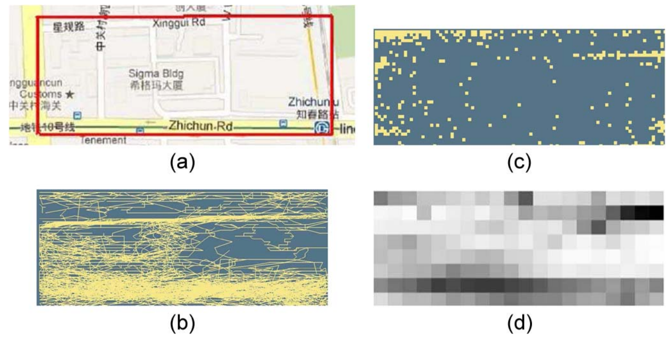

Fig. 6. Real movement traces with 612 user trajectories in Beijing.

2) **Step 2—Select one participant at a time from $\mathcal{B}$ to $\mathcal{A}$**. For each participant m in $\mathcal{B}$, his/her maximum efficiency is calculated by (14), and the most efficient participant is selected in each round and is moved from $\mathcal{B}$ to $\mathcal{A}$.
3) **Step 3—Loop**. Keep selecting participants as Step 2, until the given budget for this sensing time period can afford no more participants, or the QoI requirements of all tasks are fully satisfied.

VI. Performance Evaluation

Here, we first describe our experiment setup including the used data set, and then, we present results and discussions.

A. Experiment Setup

We evaluate the proposed scheme by using the real movement traces of ordinary citizens from Microsoft Research Asia GeoLife's data set [35]. The GeoLife project has collected 182 volunteers' trajectories in Beijing for three consecutive years. Each trajectory is marked by a sequence of time-stamped GPS readings that contain users' latitude, longitude, and altitude at a given time. We adopt the following procedures to set up our simulation platform.

1) As all traces spread in different parts of Beijing, a specific rectangular region where the traces mostly appear is needed. We stored all trajectories in a geographical MySQL database and found a $200 \times 500\ \text{m}^2$ region with high movement density, as shown in Fig. 6(a), that happens to be around the area of the Microsoft Research Asia site. We used this region as the simulation area for the considered data collection application.
2) The entire region is divided into 8×20 areas of $25 \times 25\ \text{m}^2$, i.e., $|\mathcal{L}_q| = 160, \forall q \in \mathcal{Q}$. Moreover, by setting $|\mathcal{T}_q| = 2$, the lifetime of all tasks is composed of two time slots. We considered different sensing tasks in the sensing region; the different tasks include different numbers of subregions and need different amounts of samples in each subregion. For example, for task q_1, the required amount of data in a time slot is set to 100 ($r_{lt}^{q_1} = 100, \forall l \in \mathcal{L}, t \in \mathcal{T}$), the sensing region is set to 20 subregions ($L_{q_1} = 20$); for task q_2, $r_{lt}^{q_2}$ is set to 80, and L_{q_2} is set to 30.
3) All 612 trajectories in the considered region were taken as candidate participants, i.e., $|\mathcal{M}| = 612$. Since these traces were recorded at different times, in our simulation, we simply neglect their time index. For each mobile

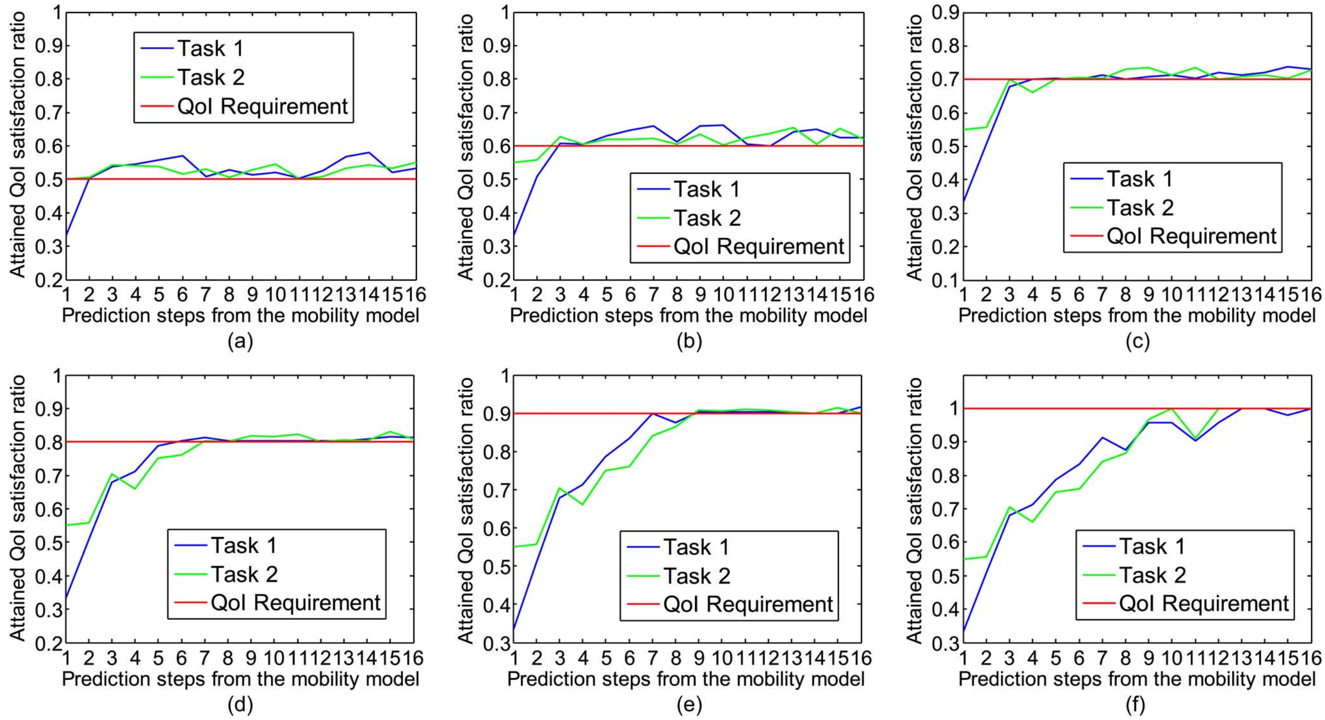

Fig. 7. Simulation results of attained QoI satisfaction ratio versus prediction steps when varying different required QoI satisfaction ratios.

user, given that the best GPS accuracy is about 5 m, we used the divided region as the mobile users' locations instead of the original GPS coordinates. The first GPS record of each trajectory that falls into the aforementioned simulation region was used as the initial location of a mobile user. Users' initial energy was randomly generated as a uniformly distributed random variable ranging from 10% to 100%.

4) To construct the location transition matrix $\underline{P}(\Delta t)$, we analyzed the adjacent movements of all 612 trajectories from one location to another. Then, we sum up the transition possibility matrix of all 612 trajectories, as shown in Fig. 7(d). Each square in the figure represents an area $l, \forall l \in \mathcal{L}$, and its gray value denotes the summed-up possibility for a participant to appear in this area from any initial location. It can also be regarded as the average amount of time that a participant spends in a specific area during the duration of simulation.

Here, we also give an example for calculating the QoI satisfaction ratio. In this example, a sensing region is divided into two subareas. A task publisher requires 28 samples for two time slots in this task q, and there are 34 samples uploaded by the participant set $\mathcal{X}$, as

$$\underline{R}^q = \begin{bmatrix} 7 & 5 \\ 9 & 7 \end{bmatrix}, \qquad \underline{O}^q(\mathcal{X}) = \begin{bmatrix} 8 & 2 \\ 19 & 5 \end{bmatrix}. \tag{16}$$

Thus, based on (8) and (9), the QoI satisfaction ratio for this task can be calculated as follows:

$$u^q(\mathcal{X}) = (\min(8/7, 1) + \min(2/5, 1) + \min(19/9, 1) + \min(5/7, 1))/4 = 0.7786. \tag{17}$$

B. Results and Discussions

As shown in Fig. 1, during the participant selection, the piece of information exchanged between the base station and smart devices only includes the properties of the remaining energy level of participants' devices and the requested incentive value, but not the actual sensory data. Moreover, the information from each participant will be reported only once. Thus, we need only a few bits for the information, as 32 bits for the remaining energy level and 32 bits for the requested incentive value in our case. To this end, the feature of low bandwidth utilization and signaling overhead makes our approach a suitable solution for MCS network. On the other hand, the personal information of each participant will not be uploaded to the central server; thus, the participants do not need to worry about the danger of privacy.

We conduct a set of simulations to explicitly evaluate the performance of our participant selection approach. First, we investigate the impact of the prediction step, which is used to calculate the probabilities of a participant moving from the current position to all other positions in the discretized sensing region, as discussed in Section III. As shown in Fig. 7, when the requested QoI satisfaction ratio g^q increases from 0.5 to 1.0, we can observe that with more mobility model prediction steps, the attained QoI satisfaction ratio grows rapidly. We also observe that this prediction step should be more than 5, if $g^q = 0.5$, but when $g^q = 1.0$ (i.e., full satisfaction is required), the prediction step of the mobility model should be more than 13. Furthermore, when the specified g^q is satisfied, it will not continue to increase by allowing more prediction steps. Therefore, the system can be operated at an appropriate state where sufficient prediction steps are enforced to receive the highest amount of QoI satisfaction given its requirement from task publishers.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

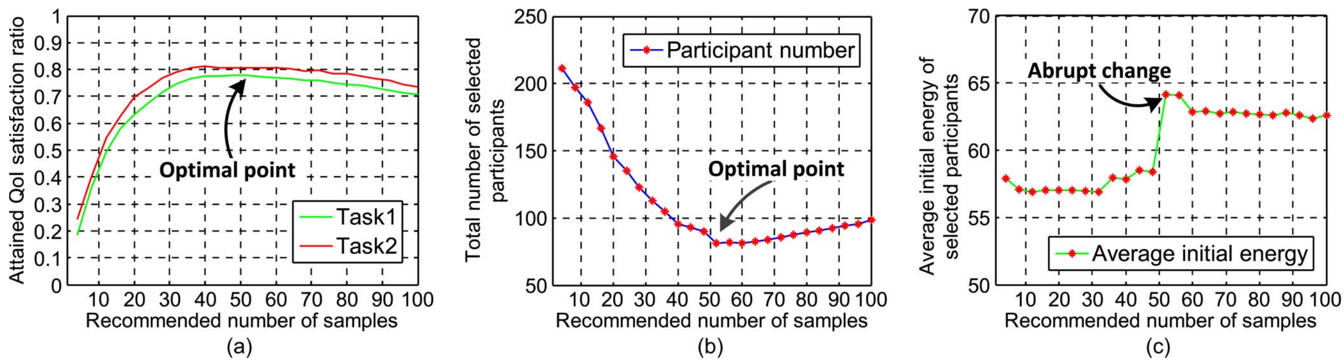

Fig. 8. Simulation results of (a) attained QoI satisfaction ratio, (b) total number of selected participants, and (c) average initial energy of selected participants when varying the recommended number of data samples.

It is worth noting that the fitted curve in (10) only represents the average user behavior/willingness in a probabilistic manner. In other words, some participants may afford more data samples while others may only allow less. Therefore, we next investigate the impact of the recommended number of data samples for all participants, as shown in Fig. 8(a). We observe that the attained QoI satisfaction ratio increases when the system requests more data samples from each participant when the volume is relatively still small, which may be acceptable by most participants. However, when the number of the demanded data samples reaches the maximum preferable amount a participant can accept (due to their limited device resources), they start to reject the sensing tasks, and this is particularly why we observe that the attained QoI satisfaction ratio starts to decrease.

Fig. 8(b) shows the impact of the recommended number of data samples on the total number of selected participants. When we increase b_m per participant, it is obvious that a fewer number of participants is required. Then, continuously increasing b_m will lead to the task rejections from some participants, who were contributing more than others. Thus, the system then tends to recruit more, but less efficient, participants although their individual contribution may not be large enough; this is why the number of selected participants increases later. Therefore, we observe an "optimal" system operating point in terms of the number of selected participants, where the operator can tune b_m to reach the desired state.

Fig. 8(c) demonstrates the impact of the recommended number of data samples on the average initial energy of those selected participants. With the increase of demanded number of samples, we observe an abrupt change of the average initial energy when 50 samples are required. This highly likely indicates the change of selected group of participants, which can be quite different from before. Combined with Fig. 8(a) and (b), it is proven that 50 is the changing point for system operations to select participants. When the recommended number of data samples is less than 50, the QoI satisfaction ratio grows rapidly, and the number of selected participants decreases; however, when the recommended number of data samples exceeds 50, both metric values go to the opposite direction. Therefore, the system should not demand a very high recommended number of data samples to participants.

Then, we conduct another set of simulations to evaluate the performance of user rejection (or human intervention). As shown in Fig. 9, we vary different numbers of required data samples, ranging from 10 to 100 units. The performance is verified against three metrics, namely, average attained QoI satisfaction ratio of all tasks, QoI loss after rejection, and attained QoI satisfaction ratio per participant. We compare our proposed approach, referred as "Proposed, before rejection," and "Proposed, after rejection," with the approach without considering user rejection, referred as "w/o considering user rejection, before rejection" and "w/o considering user rejection, after rejection." All simulation points are implemented in 100 runs with different sets of parameters, and results are averaged.

Then, we compare our proposed approach with two other approaches as benchmarks, as shown in Fig. 9(a)–(f).

1) Random selection: The central server will not consider the individual difference of participants and, thus, randomly selects a participant one by one. This process will stop when the required QoI satisfaction ratio is reached or all participants are selected.
2) Approach with equal samples: It requires the same amount of samples from each participant, and the process will also stop when the required QoI satisfaction ratio is reached or all the participants are selected. In this approach, we study the amount of samples from 10 to 100 and use the "best result" (i.e., the number of required samples (50) corresponding to the optimal point shown in Fig. 8).

As shown in Fig. 9(a), it is obvious that without considering the rejection probability, the attained QoI satisfaction ratio is high. Nevertheless, after participants perform task rejections, the achieved QoI level turns to the lowest compared with our proposed approach. We also observe that although our approach does not achieve the same QoI level as the approach that does not consider user rejection when the required amount of data samples is small, the attained QoI satisfaction ratio after user rejection is almost identical. This indicates that considering human intervention is important when calculating the overall achieved QoI level. Furthermore, we observe that when the number of required samples for all participants is 100 units, our

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

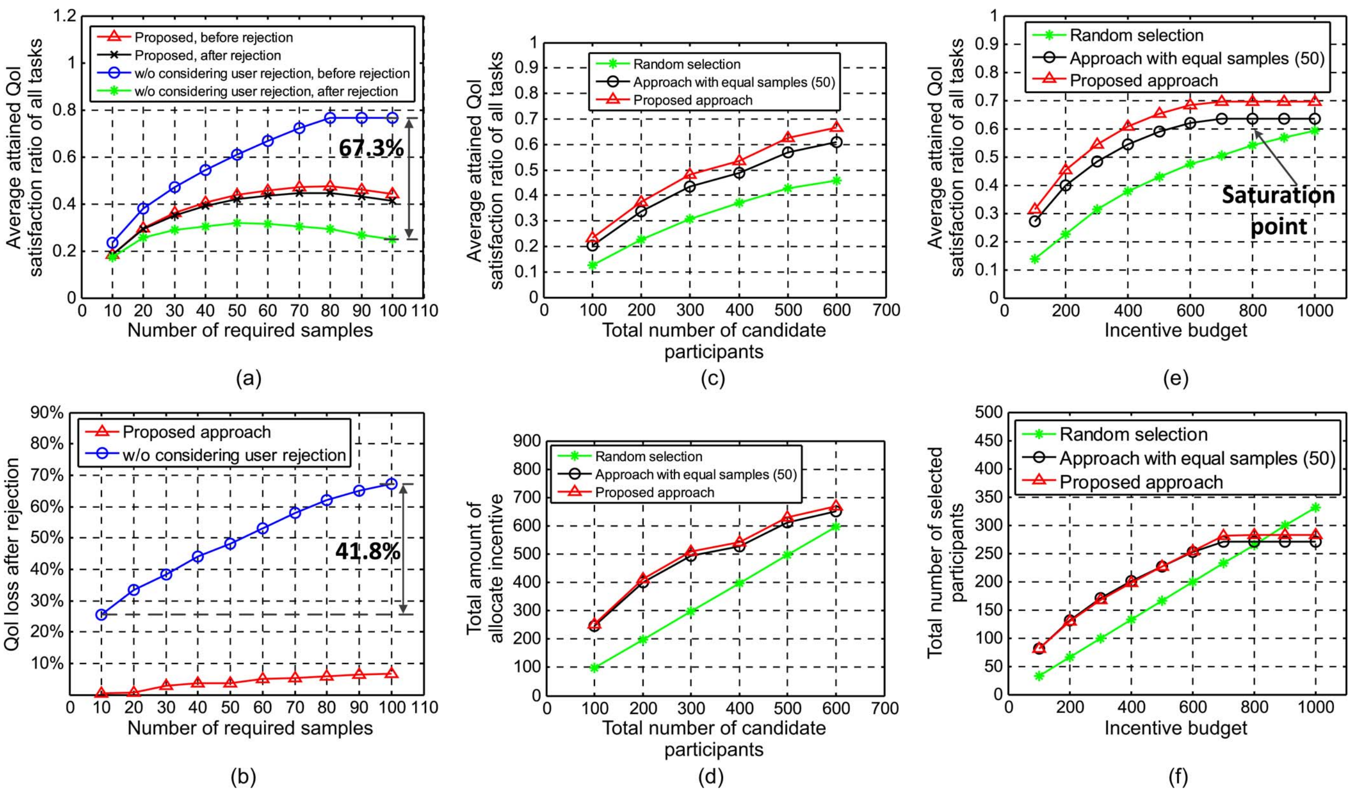

Fig. 9. Simulation results of (a) Average attained QoI satisfaction ratio of all tasks and (b) QoI loss after rejection, when varying the number of required samples; (c) Average attained QoI satisfaction ratio of all tasks and (d) total amount of allocated incentives, when varying the total number of candidate participants; (e) Average attained QoI satisfaction ratio of all tasks and (f) total number of selected participants, when varying incentive budget.

proposed approach can achieve 43.75 % more QoI satisfactions than the approach without considering user rejections.

We next examine the QoI loss, which is calculated by the difference between the attained QoI satisfaction ratios with and without considering the rejection probability. As shown in Fig. 9(b), the QoI loss of our proposed approach is less than 6.8% for different required samples, and the lowest point can reach 0.45%. This indicates that the set of selected participants is almost optimal in terms of achieving the highest QoI gain. Furthermore, we observe that the QoI loss of the approach without considering user rejection is always higher than 25.5%, and the highest point reaches up to 67.3%, with the increase of 41.8%. Therefore, our proposed approach performs more stably and reliably.

Then, we investigate the impact of the total number of candidate participants, as shown in Fig. 9(c) and (d). We randomly generated six different sets of the total number of candidate participants, ranging from 100 to 600, respectively. As shown in Fig. 9(c), the attained QoI satisfaction ratio rises with the increase of the total number of candidate participants, but the speed gradually decreases. We observe that our proposed approach can obtain 15% more QoI satisfaction ratio than the approach with equal samples when the total number of candidate participants is 100. Moreover, the QoI satisfaction ratios of these two approaches are much more than the random selection approach. However, as shown in Fig. 9(d), we observe that compared with our proposed approach, the random selection approach nearly selected all of candidate participants, and the total amount of allocated incentive for our proposed approach is almost the same with the approach with equal samples.

Finally, we investigate the impact on total incentive budget. As shown in Fig. 9(e), with the growth of total incentive budget, the average attained QoI satisfaction ratios of all three approaches keep rising. We observe that the attained QoI of our proposed approach is always much higher than the others. That is, 16% more than the approach with equal samples and 126.5% more than random selection approach, when the total incentive budget is 100. After the total incentive budget reaches a certain value, which is labeled as "saturation point" in the figure, all efficient participants calculated by our approach and the approach with equal samples are fully selected, and thus, we cannot obtain any better QoI satisfactions, but it is still higher than the one achieved by the random selection approach.

We also observe that before the incentive budget reaches the "saturation point," the number of selected participant of our proposed approach and the approach with equal samples are almost the same and are much less than the random selection approach, as shown in Fig. 9(f). Together with the result in Fig. 9(e), it is obvious that our proposed approach is more efficient than the approach with equal samples. Our proposed approach only needs a few more participants to obtain a higher QoI satisfaction ratio. Nevertheless, the random selection approach always exhausts the incentive budget; thus, the number of selected participants also keeps rising and finally exceeds other approaches.

VII. Conclusion and Future Work

In this paper, the problem of selecting an optimal set of participants in a QoI-aware and energy-efficient manner has been investigated. Specifically, a novel concept of the QoI satisfaction ratio has been introduced to quantify how much collected sensory data can satisfy a task's multidimensional QoI requirements in terms of data granularity and quantity. To build up the mathematical relationship between the recommended number of samples of a participant and his/her device's remaining energy level, a participant's sampling behavior was modeled, and task rejection probability was computed. Then, a constrained optimization problem was formulated by incorporating the above key design elements, and a suboptimal solution was proposed. Extensive simulation results, based on a real trace in Beijing, were presented to justify the effectiveness and robustness of our approach. The results are as follows.

1) A higher degree of QoI satisfaction can be obtained by using more prediction steps for the mobility model.
2) The number of data samples required from each participant is not always the more the better; an optimal value can be found by careful analysis of the participants' sampling behaviors.
3) Our proposed suboptimal solution is a very effective heuristic algorithm. Moreover, our solution can compensate the possible inefficiency of the used mobility model, which is verified by comparing with the mechanism when applying our solution even on the known trajectories.

Apart from fully considering participants' rejection probability when accepting sensing tasks, more factors can be included to improve both the quantity of quality of data. For example, a design can consider participants' reputation to only select the most reputable ones for participation, or the platform can employ a better incentive scheme to encourage participants to gather more sensory data from remote places with sparse user exposures.

References

[1] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *Proc. IEEE PERCOM Workshops*, 2014, pp. 593–598.

[2] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *J. Syst. Softw.*, vol. 84, no. 11, pp. 1928–1946, Nov. 2011.

[3] D. Mendez, M. Labrador, and K. Ramachandran, "Data interpolation for participatory sensing systems," *Pervasive Mobile Comput.*, vol. 9, no. 1, pp. 132–148, Feb. 2013.

[4] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," *IEEE Pervasive Comput.*, vol. 6030, pp. 138–155, May 2010.

[5] G. S. Tuncay, G. Benincasa, and A. Helmy, "Autonomous and distributed recruitment and data collection framework for opportunistic sensing," in *Proc. ACM MobiCom*, Oct. 2012, pp. 407–410.

[6] H. Weinschrott, F. Durr, and K. Rothermel, "Streamshaper: Coordination algorithms for participatory mobile urban sensing," in *Proc. IEEE MASS*, Nov. 2010, pp. 195–204.

[7] M. Zhong and C. G. Cassandras, "Distributed coverage control and data collection with mobile sensor networks," in *Proc. IEEE CDC*, Oct. 2010, pp. 5604–5609.

[8] H. N. Pham, B. S. Sim, and H. Y. Youn, "A novel approach for selecting the participants to collect data in participatory sensing," in *Proc. IEEE SAINT*, Jul. 2011, pp. 50–55.

[9] H. Lu, N. D. Lane, S. B. Eisenman, and A. T. Campbell, "Bubble-sensing: Binding sensing tasks to the physical world," *Pervasive Mobile Comput.*, vol. 6, no. 1, pp. 58–71, Feb. 2010.

[10] S. Gaonkar, J. Li, R. R. Choudhury, L. Cox, and A. Schmidt, "Micro-blog: Sharing and queryingcontent through mobile phones and social participation," in *Proc. ACM MobiSys*, Jun. 2008, pp. 174–186.

[11] M. Riahi, T. G. Papaioannou, I. Trummer, and K. Aberer, "Utility-driven data acquisition in participatory sensing," in *Proc. EDBT*, Mar. 2013, pp. 251–262.

[12] L. Duan *et al.*, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1701–1709.

[13] J.-S. Lee and B. Hoh, "Sell your experiences: A market mechanism based incentive for participatory sensing," in *Proc. IEEE PerCom*, Apr. 2010, pp. 60–68.

[14] I. König, A. Q. Memon, and K. David, "Energy consumption of the sensors of smartphones," in *Proc. IEEE ISWCS*, Aug. 2013, pp. 1–5.

[15] P. Baier, F. Durr, and K. Rothermel, "Psense: Reducing energy consumption in public sensing systems," in *IEEE AINA*, Mar. 2012, pp. 136–143.

[16] S. Nath, "Ace: Exploiting correlation for energy-efficient and continuous context sensing," in *Proc. ACM MobiSys*, Jun. 2012, pp. 29–42.

[17] X. Sheng, J. Tang, and W. Zhang, "Energy-efficient collaborative sensing with mobile phones," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1916–1924.

[18] J. A. Burke *et al.*, "Participatory sensing," in *Proc. ACM SenSys*, Nov. 2006, pp. 1–5.

[19] P. Dutta *et al.*, "Common sense: Participatory urban sensing using a network of handheld air quality monitors," in *Proc. ACM SenSys*, Nov. 2009, pp. 349–350.

[20] R. K. Rana, C. T. Chou, S. S. Kanhere, N. Bulusu, and W. Hu, "Ear-phone: An end-to-end participatory urban noise mapping system," in *Proc. ACM/IEEE IPSN*, Apr. 2010, pp. 105–116.

[21] E. Kanjo, "Noisespy: A real-time mobile phone platform for urban noise monitoring and mapping," *Mobile Netw. Appl.*, vol. 15, no. 4, pp. 562–574, Aug. 2010.

[22] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Demo: Medusa: A programming framework for crowd-sensing applications," in *Proc. ACM MobiSys*, Jun. 2012, pp. 481–482.

[23] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.

[24] G. Smith, R. Wieser, J. Goulding, and D. Barrack, "A refined limit on the predictability of human mobility," in *Proc. IEEE PerCom*, 2014, pp. 88–94.

[25] D. Zhang, M. Chen, M. Guizani, and H. Xiong, "Mobility prediction in telecom cloud using mobile calls," *IEEE Wireless Commun.*, vol. 21, no. 1, pp. 26–32, Feb. 2014.

[26] S. Gambs, M.-O. Killijian, and M. N. del Prado Cortez, "Next place prediction using mobility Markov chains," in *Proc. 1st Workshop Meas., Privacy, Mobility*, 2012, Art. ID. 3.

[27] Z. Ruan, E. Ngai, and J. Liu, "Wireless sensor network deployment in mobile phones assisted environment," in *Proc. IEEE IWQoS*, 2010, pp. 1–9.

[28] Y. Wu, Y. Zhu, and B. Li, "Infrastructure-assisted routing in vehicular networks," in *Proc. IEEE INFOCOM*, 2012, pp. 1485–1493.

[29] D. Zhang, L. Wang, H. Xiong, and B. Guo, "4W1H in mobile crowd sensing," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 42–48, Aug. 2014.

[30] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. C. Leung, "Cap: Community activity prediction based on big data analysis," *IEEE Netw.*, vol. 28, no. 4, pp. 52–57, Jul./Aug. 2014.

[31] G. Cardone, A. Cirri, A. Corradi, and L. Foschini, "The participact mobile crowd sensing living lab: The testbed for smart cities," *IEEE Commun. Mag.*, vol. 52, no. 10, pp. 78–85, 2014.

[32] V. Pankratius, F. Lind, A. Coster, P. Erickson, and J. Semeter, "Mobile crowd sensing in space weather monitoring: The Mahali project," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 22–28, Aug. 2014.

[33] X. Lu, D. Li, B. Xu, W. Chen, and Z. Ding, "Minimum cost collaborative sensing network with mobile phones," in *Proc. IEEE ICC*, 2013, pp. 1816–1820.

[34] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.

[35] Y. Zheng, X. Xie, and W. Ma, "Geolife: A collaborative social networking service among user, location and trajectory," *IEEE Data Eng. Bull.*, vol. 33, no. 2, pp. 32–40, 2010.

Chi Harold Liu (M'10) received the B.Eng. degree from Tsinghua University, Beijing, China, and the Ph.D. degree from Imperial College London, London, U.K.

He is a Full Professor with the School of Software, Beijing Institute of Technology, Beijing. He is also the Director of the IBM Mainframe Excellence Center (Beijing), the IBM Big Data Technology Center, and the National Laboratory of Data Intelligence for China Light Industry. Before moving to academia, he joined IBM Research—China as a Staff Researcher and a Project Manager, after working as a Postdoctoral Researcher with Deutsche Telekom Laboratories, Germany, and a Visiting Scholar with the IBM T. J. Watson Research Center, USA. He has published more than 60 prestigious conference and journal papers and is the holder of more than ten EU/U.S./China patents. His current research interests include the Internet of things (IoT), big data analytics, mobile computing, and wireless ad hoc, sensor, and mesh networks.

Dr. Liu received the Distinguished Young Scholar Award in 2013, the IBM First Plateau Invention Achievement Award in 2012, and the IBM First Patent Application Award in 2011 and was interviewed by EEWeb.com as the Featured Engineer in 2011. He serves as an Editor for the KSII Transactions on Internet and Information Systems and a Book Editor for four books published by Taylor & Francis Group, USA. He has also served as the General Chair of the IEEE SECON'13 Workshop on Internet-of-Things Networking and Control, the IEEE WCNC'12 Workshop on IoT Enabling Technologies, and the ACM UbiComp'11 Workshop on Networking and Object Memories for IoT. He served as a Consultant to the Asian Development Bank, Bain & Company, and KPMG, USA, and a Peer Reviewer for the Qatar National Research Foundation and the National Science Foundation, China. He is a member of the Association for Computing Machinery.

Bo Zhang received the M.S. degree from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2010, where he is currently working toward the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology.

Prior to studying at BUPT, he was with the R&D department of H3C Technologies Co., Ltd. His research interests include participatory sensing and the Internet-of-things.

Xin Su (M'02) received the M.S. and Ph.D. degrees in electronic engineering from the University of Electronic Science and Technology of China, Chengdu, China, in 1996 and 1999, respectively.

From January 2000 to October 2001, he was a Research Fellow with the Department of Electronic Engineering, Tsinghua University, Beijing, China, where he worked on software radio theory and technology and applications for personal communications. Since November 2001, he has been with the School of Information Science and Technology, Tsinghua University, where he is currently a Professor with the Research Institute of Information Technology and the Vice Director of the Wireless and Mobile Communications Technology R&D Center. As a Faculty Member and Graduate Supervisor, he has been teaching courses in wireless and mobile communications. His research interests include wireless and mobile networks, broadband wireless access, software-defined radio, and wireless multimedia communications.

Dr. Su is the Chairman of the International Mobile Telecommunications Advanced (IMT-Advanced) Technology Working Group, Ministry of Industry and Information Technology of China and the Vice Chairman of the Innovative Wireless Technology Working Group, China Communications Standards Association.

Jian Ma received the B.Sc. and M.Sc. degrees from Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 1982 and 1987, respectively, and the Ph.D. degree from the Department of Electronics Engineering, Helsinki University of Technology, Espoo, Finland, in 1994.

He is a Principal Member of Research Staff with the Nokia Research Center, Beijing. He was a Guest Professor in computer science with BUPT in 2002, ICT-CAS in 2003, and CCCE-GUCAS in 2006 and is currently supervising about one and a half dozen Ph.D. and M.Sc. students at BUPT, CCCE-GUCAS, ICT-CAS, and Tsinghua University, Beijing. He authored more than 160 conference and journal papers and four books and book chapters.

Dr. Ma has been involved in few professional activities, currently serving as a Council Member of the Beijing Communication Institute (a senior consultant club in communication industry and academia) and a Vice Chair of the Sensor Network Technical Committee, China Computer Federation. He has also been serving as a Cochair and member of organization committees or program committees in several leading conferences, such as ICC, AINA, PIMRC, ICDCS, WWW, and ChinaCom.

Wendong Wang (M'05) received the Bachelor's and Master's degrees from Beijing University of Posts and Telecommunications, Beijing, China, in 1985 and 1991, respectively.

He is currently a Full Professor with Beijing University of Posts and Telecommunications and on the Assessment Panel of the "National Natural Science Foundation" program and the "National High Technology Research and Development Program of China." He has published more than 200 papers in various journals and conference proceedings and is the holder of 14 U.S./China patents. His current research interests are the next-generation network architecture; Internet-of-things; participatory sensing; wireless ad hoc, sensor, and mesh networks; and mobile Internet.

Mr. Wang is a member of the Association for Computing Machinery.

Kin K. Leung (F'01) received the B.S. degree from The Chinese University of Hong Kong, Shatin, Hong Kong, in 1980 and the M.S. and Ph.D. degrees in computer science from the University of California, Los Angeles, CA, USA, in 1982 and 1985, respectively.

He joined AT&T Bell Labs in 1986 and worked for its successor companies, AT&T Labs and Bell Labs of Lucent Technologies, until 2004. Since then, he has been the Tanaka Chair Professor in Internet technology with Imperial College London, London, U.K. His research interests include network resource allocation, MAC protocol, TCP/IP protocol, mobility management, network architecture, real-time applications, and teletraffic issues for wireless IP networks, sensor, and ad hoc networks. He is also interested in a wide variety of wireless technologies, including IEEE 802.11, 802.16, and 3G and future-generation cellular networks.