# Budget Feasible Online Incentive Mechanisms for Crowdsourcing Tasks Truthfully

Dong Zhao, Xiang-Yang Li, *Senior Member, IEEE,* and Huadong Ma, *Member, IEEE*

*Abstract*—Mobile Crowd Sensing (MCS) is a new paradigm which takes advantage of pervasive mobile devices to efficiently collect data, enabling numerous novel applications. To achieve good service quality for an MCS application, incentive mechanisms are necessary to attract more user participation. Most of existing mechanisms apply only for the *offline* scenario where all users report their strategic types in advance. On the contrary, we focus on a more realistic scenario where users arrive one by one *online* in a random order. Based on the *online auction* model, we investigate the problem that users submit their private types to the crowdsourcer when arrive, and the crowdsourcer aims at selecting a subset of users before a specified deadline for maximizing the value of services (assumed to be a non-negative monotone submodular function) provided by selected users under a budget constraint. We design two *online mechanisms*, *OMZ* and *OMG*, satisfying the *computational efficiency*, *individual rationality*, *budget feasibility*, *truthfulness*, *consumer sovereignty* and *constant competitiveness* under the zero arrival-departure interval case and a more general case, respectively. Through extensive simulations, we evaluate the performance and validate the theoretical properties of our online mechanisms.

*Index Terms*—Crowdsourcing, incentive mechanism design, online auction.

## I. Introduction

Crowdsourcing is a distributed problem-solving model in which a crowd of undefined size is engaged to solve a complex problem through an open call [1]. Nowadays, the proliferation of mobile devices (e.g., smartphones, wearable devices, in-vehicle sensing devices) provides a new opportunity for extending existing web-based crowdsourcing applications to a larger contributing crowd, making contribution easier and omnipresent. Furthermore, today's smartphones are programmable and come with a rich set of cheap powerful embedded sensors, such as GPS, WiFi/3G/4G interfaces, accelerometer, digital compass, gyroscope, microphone, and camera. The great potential of mobile sensing offers a variety of novel, efficient ways to collect data, enabling numerous *Mobile Crowd Sensing* (MCS) applications, such as Sensorly [2] for constructing cellular/WiFi network coverage maps, SignalGuru [3], Nericell [4] and VTrack [5] for providing traffic information, Ear-Phone [6] and NoiseTube [7] for making noise maps. For more details on MCS, we refer interested readers to several survey papers [1], [8], [9].

D. Zhao and H. Ma are with the Beijing Key Lab of Intelligent Telecomm. Software and Multimedia, Beijing University of Posts and Telecomm., Beijing, 100876, China. E-mail: {dzhao, mhd}@bupt.edu.cn.

X.-Y. Li is with the Department of Computer Science, Illinois Institute of Technology, Chicago, IL, USA. E-mail: xli@cs.iit.edu.

This work was done during Dong Zhao's visit at the Illinois Institute of Technology.

Adequate user participation is one of the most critical factors determining whether an MCS application can achieve good service quality. Most of the current MCS applications [2]–[7] are based on voluntary participation. While participating in an MCS campaign, mobile users consume their own resources such as battery and computing power, and expose their locations with potential privacy threats. Thus, incentive mechanisms are necessary to provide participants with enough rewards for their participation costs. At present, a lot of research [10]–[16] focuses on incentive mechanism design for MCS applications. Most of existing mechanisms apply only for the *offline* scenario in which all of participating users report their strategic types, including the tasks they can complete and the bids, to the crowdsourcer (campaign organizer) in advance, and then the crowdsourcer selects a subset of users after collecting the information of all users to maximize his/her utility (e.g., the total value of all tasks that can be completed by selected users).

In practice, however, users always arrive one by one *online* in a random order and user availability changes over time. Therefore, an *online incentive mechanism* is necessary to make irrevocable decisions on whether to accept a user's task and bid, based solely on the information of users arriving before the present moment, without knowing future information.

In this paper we consider a general problem: the crowdsourcer aims at selecting a subset of users before a specified deadline, so that the value of services provided by selected users is maximized under the condition that the total payment to these users does not exceed a budget constraint. Specially, we investigate the case where the value function of selected users is non-negative monotone submodular. This covers many realistic scenarios. For example, many MCS applications [2]–[7] aim at selecting users to collect sensing data so that a given region can be covered before a specified deadline, where the coverage function is typically non-negative monotone submodular. We further assume that the cost and arrival/departure time of each user are private and only known to itself. Users are assumed to be game-theoretic and seek to make strategy (possibly report an untruthful cost or arrival/departure time) to maximize their individual utility in equilibrium. Thus, the problem of selecting crowdsourcing users while maximizing the value can be modeled as an *online auction*.

Our objective is to design online mechanisms satisfying six desirable properties: *computational efficiency*, *individual rationality*, *budget feasibility*, *truthfulness*, *consumer sovereignty* and *constant competitiveness*. Informally, *computational efficiency* ensures the mechanism can run in real time, *individual rationality* ensures each participating user has a non-negative

utility, *budget feasibility* ensures the crowdsourcer's budget constraint is not violated, *truthfulness* ensures the participating users report their true costs (*cost-truthfulness*) and arrival/departure times (*time-truthfulness*), *consumer sovereignty* ensures each participating user has a chance to win the auction, and *constant competitiveness* guarantees that the mechanism performs close to the optimal solution in the offline scenario where all users' information are known a priori. Although several recent studies also investigate online mechanisms for crowdsourcing markets [17]–[19], they focus on different models or goals, or fail to consider some important properties.

The main idea behind our online mechanisms is to adopt a multiple-stage sampling-accepting process. At every stage the mechanism allocates tasks to a user only if his/her marginal density is not less than a certain density threshold that has been computed using previous users' information, and the budget allocated for the current stage has not been exhausted. Meanwhile, the user obtains a bid-independent payment. The density threshold is computed in a manner that guarantees desirable performance properties of the mechanism. We first consider the *zero arrival-departure interval* case where the arrival time of each user equals to his/her departure time (Section III). In this case, achieving time-truthfulness is trivial. We present an online mechanism *OMZ* satisfying all desirable properties under this special case without considering the time-truthfulness. Then we revise the *OMZ* mechanism, and present another online mechanism *OMG* satisfying all desirable properties under a more *general* case (Section IV).

The remainder of this paper is organized as follows. In Section II we describe the MCS system model, and formulate the problem as an online auction. We then present two online mechanisms, *OMZ* and *OMG*, satisfying all desirable properties under the *zero arrival-departure interval* case and the *general* case in Section III and IV, respectively. Performance evaluations are presented in Section V. We review the related work in Section VII, and conclude this paper in Section VIII.

## II. SYSTEM MODEL AND PROBLEM FORMULATION

We use Fig. 1 to illustrate an MCS system. The system consists of a *crowdsourcer*, which resides in the cloud and consists of multiple sensing servers, and many smartphone *users*, which are connected to the cloud by cellular networks (e.g., GSM/3G/4G) or Wi-Fi connections. The crowdsourcer first publicizes an MCS campaign in a *Region of Interest (RoI)*, aiming at finding some users to complete a set of tasks $\Gamma = \{\tau_1, \tau_2, \ldots, \tau_m\}$ in the RoI before a specified deadline $T$. Assume that a crowd of mobile users $\mathcal{U} = \{1, 2, \ldots, n\}$ interested in participating in the campaign arrive online in a random order, where $n$ is unknown. Each user $i$ has an arrival time $a_i \in \{1, \ldots, T\}$, a departure time $d_i \in \{1, \ldots, T\}$, $d_i \geq a_i$, and a subset of tasks $\Gamma_i \subseteq \Gamma$ he/she can complete within this time interval. Meanwhile, user $i$ also has an associated cost $c_i \in \mathbb{R}_+$ for performing sensing tasks. All information constitutes the *type* of user $i$, $\theta_i = (a_i, d_i, \Gamma_i, c_i)$. In this paper we consider two models with respect to the distribution of users:

- **The i.i.d. model:** The costs and values of users are i.i.d. sampled from some *unknown* distributions.
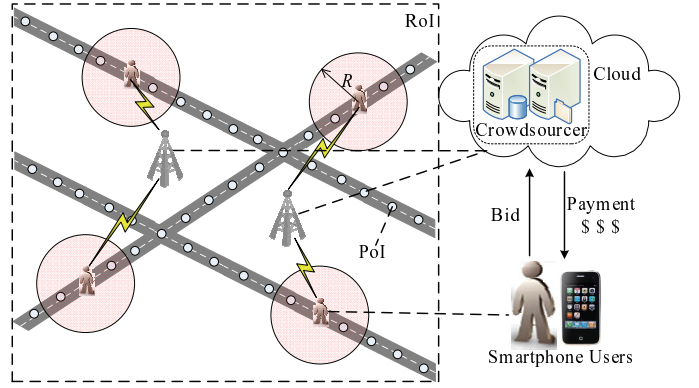


Fig. 1. Illustration of a mobile crowd sensing system.

- **The secretary model:** An adversary gets to decide on the costs and values of users, but not on the *order* in which they are presented to the crowdsourcer.

In fact, the i.i.d. model is a special case of the secretary model, since the sequence can be determined by first picking a multiset of costs or values from the (unknown) distribution, and then permuting them randomly. Note that these two models are different from the *oblivious adversarial model*, where an adversary chooses a *worst-case* input stream including the users' costs, values and their arrival orders.

We model the interactive process between the crowdsourcer and users as an *online auction*. Each user expects a *payment* in return for his/her service. Therefore, he/she makes a price that he/she expects, called *bid*, for selling his/her sensing data. When a user arrives, the crowdsourcer must decide whether to buy the service of this user, and if so, at what price, before he/she departs. Assume that the crowdsourcer has a budget constraint $B$ indicating the maximum value that he/she is willing to pay. Therefore, the crowdsourcer always expects to obtain the maximum value from the selected users' services under the budget constraint.

Users are assumed to be game-theoretic and seek to make *strategy* to maximize their individual utility in equilibrium. Note that the arrival/departure time and cost of user $i$ are private and only known to himself/herself. Only the task set $\Gamma_i$ must be true since the crowdsourcer can identify whether the announced tasks are performed [*]. In other words, user $i$ may misreport all information about his/her type except for $\Gamma_i$. The budget and value function of the crowdsourcer are common knowledge. Although we do not require a user to declare his/her departure time until the moment of his/her departure, we find it convenient to analyze our auctions as direct-revelation mechanisms (DRMs) [21]. The strategy space in an online DRM allows a user to declare some possibly untruthful type $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \Gamma_i, b_i)$, subject to $a_i \leq \hat{a}_i \leq \hat{d}_i \leq d_i$. Note that we assume that a user cannot announce an earlier arrival time or a later departure time than his/her true arrival/departure time. This assumption is justified if the user's presence can be directly verified, or (in the case of "no early arrivals") if we

---

[*]If the user fails to complete the tasks, or fails to meet the time or quality requirement, the crowdsourcer can give no payment or less payment, or records his/her reputation for assisting in future decision [20].

think of $a_i$ as the time at which the user first becomes aware of the existence of the auction or of his own desire to participate in it [22].

In order to obtain the required service, the crowdsourcer needs to design an *online mechanism* $\mathcal{M} = (f, p)$ consisting of an *allocation* function $f$ and a *payment* function $p$. For any *strategy sequence* $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$, the allocation function $f(\hat{\theta})$ computes an allocation of tasks for a selected subset of users $\mathcal{S} \subseteq \mathcal{U}$, and the payment function $p(\hat{\theta})$ returns a vector $(p_1(\hat{\theta}), \ldots, p_n(\hat{\theta}))$ of payments to the users. Note that, the crowdsourcer, when presented with the strategy $\hat{\theta}_i$ of user $i$, must decide whether to accept user $i$ at what price ($p_i$) before the time step $\hat{d}_i$.

The *utility* of user $i$ is

$$u_i = \begin{cases} p_i - c_i, & \text{if} \quad i \in \mathcal{S}; \\ 0, & \text{otherwise.} \end{cases}$$

Let $V(\mathcal{S})$ denote the *value* function of the crowdsourcer over the selected subset of users $\mathcal{S}$. The crowdsourcer expects to obtain the maximum value from the selected users' services under the budget constraint, i.e.,

**Maximize** $V(\mathcal{S})$ **subject to** $\sum_{i \in \mathcal{S}} p_i \leq B.$

In this paper, we focus on the case where $V(\mathcal{S})$ is non-negative monotone submodular. This covers many realistic scenarios.

**Definition 1** (Monotone Submodular Function). *Let $\Omega$ be a finite set. For any $X \subseteq Y \subseteq \Omega$ and $x \in \Omega \backslash Y$, a function $f : 2^\Omega \mapsto \mathbb{R}$ is called submodular if and only if*

$$f(X \cup \{x\}) - f(X) \geq f(Y \cup \{x\}) - f(Y),$$

*and it is monotone (increasing) if and only if $f(X) \leq f(Y)$, where $2^\Omega$ denotes the power set of $\Omega$, and $\mathbb{R}$ denotes the set of reals.*

Application Scenario Illustration: As illustrated in Fig. 1, we consider the scenario where the crowdsourcer expects to obtain the sensing data covering all roads in an RoI. For example, in Sensorly [2] or Ear-Phone [6], the crowdsourcer allocates tasks to the smartphone users for constructing sensing maps of Wi-Fi signals or environmental noises. For convenience of calculations, we divide each road in the RoI into multiple discrete *Points of Interest (PoIs)*, and the objective of the crowdsourcer is equivalent to obtaining the sensing data covering all PoIs before $T$. The set of PoIs is denoted by $\Gamma = \{\tau_1, \tau_2, \ldots, \tau_m\}$. Assume that each sensor follows a geometric disk sensing model with sensing range $R$, which means if user $i$ senses at a location $L_i$ and obtain a reading, then any PoI within the disk with the origin at $L_i$ and a radius of $R$ has been covered once. The set of PoIs covered by user $i$ is denoted by $\Gamma_i \subseteq \Gamma$, which means the sensing tasks that user $i$ can complete. Without loss of generality, assume that each PoI $\tau_j$ has a coverage requirement $r_j \in \mathbb{Z}_+$ indicating how many times it requires to be sensed at most. The *value* of the selected users to the crowdsourcer is:

$$V(\mathcal{S}) = \sum_{j=1}^{m} \min\{r_j, \sum_{i \in \mathcal{S}} v_{i,j}\},$$

where $v_{i,j}$ equals to 1 if $\tau_j \in \Gamma_i$, and 0 otherwise.

**Lemma 1.** *The value function $V(\mathcal{S})$ is monotone submodular.*

The proof of Lemma 1 is given in Appendix A.

Our objective is to design an online mechanism satisfying the following six desirable properties:

- **Computational Efficiency:** A mechanism is *computationally efficient* if both the allocation and payment can be computed in polynomial time as each user arrives.
- **Individual Rationality:** Each participating user will have a non-negative utility: $u_i \geq 0$, if he/she reports the true cost and arrival/departure time.
- **Budget Feasibility:** We require the mechanism to be *budget feasible*: $\sum_{i \in \mathcal{S}} p_i \leq B$.
- **Truthfulness:** A mechanism is *cost-* and *time-truthful* (or simply called *truthful*, or *incentive compatible* or *strategyproof*) if reporting the true cost and arrival/departure time is a *dominant strategy* for all users [†]. In other words, no user can improve his/her utility by submitting a false cost, or arrival/departure time, no matter what others submit.
- **Consumer Sovereignty:** The mechanism cannot arbitrarily exclude a user; the user will be selected by the crowdsourcer and obtain a payment if only his/her bid is sufficiently low while others are fixed.
- **Competitiveness:** The goal of the mechanism is to maximize the value of the crowdsourcer. To quantify the performance of the mechanism we compare its solution with the *optimal solution*: the solution obtainable in the offline scenario where the crowdsourcer has full knowledge about users' types. A mechanism is $O(g(n))$-*competitive* if the ratio between the online solution and the optimal solution is $O(g(n))$. Ideally, we would like our mechanism to be $O(1)$-*competitive*.

The importance of the first three properties is obvious, because they together guarantee that the mechanism can be implemented in real time and satisfy the basic requirements of both the crowdsourcer and users. In addition, the last three properties are indispensable for guaranteeing that the mechanism has high performance and robustness. The truthfulness aims at eliminating the fear of market manipulation and the overhead of strategizing over others for the participating users. The consumer sovereignty aims at guaranteeing that each participating user has a chance to win the auction and obtain a payment, otherwise it will hinder the users' completion or even result in task starvation. Besides, if some users are guaranteed not to win the auction, then being truthful or not will have the same outcome. For this reason, the property satisfying both the consumer sovereignty and the truthfulness is also called *strong truthfulness* by Hajiaghayi et al. [21]. Later we will show that satisfying consumer sovereignty is not trivial in the online scenario, which is in contrast to the offline scenario. Finally, we expect that our mechanism has a constant competitiveness under both the *i.i.d.* model and the *secretary* model. Note that no constant-competitive auction is possible under the *oblivious adversarial* model [23].

[†]It also implies the "dominant strategy equilibrium" in the mechanism [21].

TABLE I
FREQUENTLY USED NOTATIONS.

| Notation | Description |
|---|---|
| $\mathcal{U}, n, i$ | set of users, number of users, and one user |
| $\Gamma, m, \tau_j$ | set of tasks, number of tasks, and one task |
| $B, B'$ | budget constraint and stage-budget |
| $T, T', t$ | deadline, end time step of each stage, and each time step |
| $a_i, \hat{a}_i$ | true arrival time and strategic arrival time of user $i$ |
| $d_i, \hat{d}_i$ | true departure time and strategic departure time of user $i$ |
| $\Gamma_i$ | set of user $i$'s tasks |
| $c_i, b_i$ | true cost and bid of user $i$ |
| $\theta_i, \hat{\theta}_i$ | true type and strategy of user $i$ |
| $\mathcal{S}, \mathcal{S}'$ | set of selected users and sample set |
| $p_i, u_i$ | payment and utility of user $i$ |
| $V(\mathcal{S})$ | value function of the crowdsourcer over $\mathcal{S}$ |
| $V_i(\mathcal{S})$ | marginal value of user $i$ over $\mathcal{S}$ |
| $\rho^*$ | density threshold |
| $\delta$ | parameter used for computing the density threshold |
| $\omega$ | parameter assumed on users' value |

Table I lists frequently used notations.

## III. ONLINE MECHANISM UNDER ZERO ARRIVAL-DEPARTURE INTERVAL CASE

In this section, we consider a special case where the arrival time of each user equals to his/her departure time. In this case, each user is impatient since the decision must be made immediately once he/she arrives. Note that achieving time-truthfulness is trivial in this case. It is because that any user has no incentive to report a later arrival time or an earlier departure time than his/her true arrival/departure time, since the user cannot perform any sensing task or obtain a payment after he/she departs. In this section, we present an online mechanism *OMZ* satisfying all desirable properties under this *zero arrival-departure interval* case, without considering the time-truthfulness. Then, in Section IV we revise this mechanism and prove the revised one satisfies all desirable properties including the time-truthfulness under the general case without *zero arrival-departure interval* assumption. To facilitate understanding, in this section it is also assumed that no two users have the same arrival time. Note that this assumption can also be easily removed according to the revised mechanism in Section IV.

### A. Mechanism Design

An online mechanism needs to overcome several nontrivial challenges: first, the users' costs are unknown and need to be reported in a truthful manner; second, the total payment cannot exceed the crowdsourcer's budget; finally, the mechanism needs to cope with the online arrival of users. Previous solutions of online auctions and related problems [22], [24] always achieve desirable outcomes in online settings via a two-stage sampling-accepting process: the first batch of users is rejected and used as the sample which enables making an informed decision on whether accepting the rest of users. However, these solutions cannot guarantee the *consumer sovereignty*, since the first batch of users has no chance to win the auction no matter how low his/her cost is. It can lead to undesirable effects in our problem: automatically rejecting the first batch of users encourages users to arrive late; in other words, those users
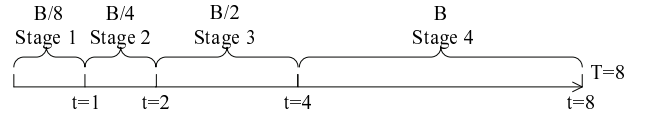


Fig. 2. Illustration of a multiple-stage sampling-accepting process when $T = 8$.

arriving early have no incentive to report their bids, which may hinder the users' competition or even result in task starvation.

To address the above challenges, we design our online mechanism, *OMZ*, based on a *multiple-stage sampling-accepting* process. The mechanism dynamically increases the sample size and learns a *density threshold* used for future decision, while increasing the *stage-budget* it uses for allocation at various stages. The rationale behind this idea contains two points. First, each stage is an accepting process as well as a sampling process ready for the next stage, so users are not automatically rejected during the sampling process. Second, it adopts an "incremental learning" process. At first, the sample size is small, so we need to learn the density threshold frequently. As the sample size increases, a more accurate density threshold can be learned, so we can decrease the learning frequency gradually. The learned density threshold will be accurate enough when about a half of users have arrived (to be proved later in Lemma 7 and Lemma 10), so at that time, we can stop the learning process.

The whole process is illustrated in Algorithm 1. First, we divide all of $T$ time steps into $(\lfloor \log_2 T \rfloor + 1)$ stages: $\{1, 2, \ldots, \lfloor \log_2 T \rfloor, \lfloor \log_2 T \rfloor + 1\}$. The stage $i$ ends at time step $T' = \lfloor 2^{i-1} T / 2^{\lfloor \log_2 T \rfloor} \rfloor$. Correspondingly, the stage-budget for the $i$-th stage is allocated as $B' = 2^{i-1} B / 2^{\lfloor \log_2 T \rfloor}$. Fig. 2 is an illustration when $T = 8$. When a stage is over, we add all users who have arrived into the sample set $\mathcal{S}'$, and compute a density threshold $\rho^*$ according to the information of samples and the allocated stage-budget $B'$. This density threshold is computed by calling the **GetDensityThreshold** algorithm (to be elaborated later), and used for making decision at the next stage. Specially, when the last stage $i = \lfloor \log_2 T \rfloor + 1$ comes, the density threshold has been computed according to the information of all users arriving before time step $\lfloor T/2 \rfloor$, and the allocated stage-budget $B/2$.

---

**Algorithm 1:** Online Mechanism under Zero Arrival-departure Interval Case (*OMZ*)

**Input**: Budget constraint $B$, deadline $T$

1   $(t, T', B', \mathcal{S}', \rho^*, \mathcal{S}) \leftarrow (1, \frac{T}{2^{\lfloor \log_2 T \rfloor}}, \frac{B}{2^{\lfloor \log_2 T \rfloor}}, \emptyset, \epsilon, \emptyset)$;
2   **while** $t \leq T$ **do**
3     **if** *there is a user $i$ arriving at time step $t$* **then**
4       **if** $b_i \leq V_i(\mathcal{S})/\rho^* \leq B' - \sum_{j \in \mathcal{S}} p_j$ **then**
5        $p_i \leftarrow V_i(\mathcal{S})/\rho^*$; $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$;
6       **else** $p_i \leftarrow 0$;
7       $\mathcal{S}' \leftarrow \mathcal{S}' \cup \{i\}$;
8     **end**
9     **if** $t = \lfloor T' \rfloor$ **then**
10       $\rho^* \leftarrow$ **GetDensityThreshold**$(B', \mathcal{S}')$;
11       $T' \leftarrow 2T'$; $B' \leftarrow 2B'$;
12     **end**
13     $t \leftarrow t + 1$;
14 **end**

Given a set of selected users $\mathcal{S}$, the *marginal value* of user $i \notin \mathcal{S}$ is $V_i(\mathcal{S}) = V(\mathcal{S} \cup \{i\}) - V(\mathcal{S})$, and his/her *marginal density* is $V_i(\mathcal{S})/b_i$. When a new user $i$ arrives, the mechanism allocates tasks to him/her as long as his/her marginal density is not less than the current density threshold $\rho^*$, and the allocated stage-budget $B'$ has not been exhausted. Meanwhile, we give user $i$ a payment

$$p_i = V_i(\mathcal{S})/\rho^*,$$

and add this user to the set of selected users $\mathcal{S}$. To start the mechanism, we initially set a small density threshold $\epsilon$, which is used for making decision at the first stage.

Since each stage maintains a common density threshold, it is natural to adopt a *proportional share* allocation rule to compute the density threshold from the sample set $\mathcal{S}'$ and the allocated stage-budget $B'$. As illustrated in Algorithm 2, the computation process adopts a greedy strategy. Users are sorted according to their increasing marginal densities. In this sorting the $(i+1)$-th user is the user $j$ such that $V_j(\mathcal{J}_i)/b_j$ is maximized over $\mathcal{S}' \backslash \mathcal{J}_i$, where $\mathcal{J}_i = \{1, 2, \ldots, i\}$ and $\mathcal{J}_0 = \emptyset$. Considering the submodularity of $V$, this sorting implies that:

$$\frac{V_1(\mathcal{J}_0)}{b_1} \geq \frac{V_2(\mathcal{J}_1)}{b_2} \geq \cdots \geq \frac{V_{|\mathcal{S}'|}(\mathcal{J}_{|\mathcal{S}'|-1})}{b_{|\mathcal{S}'|}}.$$

Find the largest $k$ such that $b_k \leq \frac{V_k(\mathcal{J}_{k-1})B}{V(\mathcal{J}_k)}$. The set of selected users is $\mathcal{J}_k = \{1, 2, \ldots, k\}$. Finally, we set the density threshold to be $\frac{V(\mathcal{J}_k)}{\delta B'}$. Here we set $\delta > 1$ to obtain a slight underestimate of the density threshold for guaranteeing enough users selected and avoiding the waste of budget. Later we will fix the value of $\delta$ elaborately to enable the mechanism achieving a constant competitive ratio.

---

**Algorithm 2:** GetDensityThreshold

**Input**: Stage-budget $B'$, sample set $\mathcal{S}'$
1   $\mathcal{J} \leftarrow \emptyset$; $i \leftarrow \arg\max_{j \in \mathcal{S}'}(V_j(\mathcal{J})/b_j)$;
2   **while** $b_i \leq \frac{V_i(\mathcal{J})B'}{V(\mathcal{J} \cup \{i\})}$ **do**
3      $\mathcal{J} \leftarrow \mathcal{J} \cup \{i\}$;
4      $i \leftarrow \arg\max_{j \in \mathcal{S}' \backslash \mathcal{J}}(V_j(\mathcal{J})/b_j)$;
5   **end**
6   $\rho \leftarrow V(\mathcal{J})/B'$;
7   **return** $\rho/\delta$;

---

In the following, we use an example to illustrate how the *OMZ* mechanism works.

**Example 1.** *Consider a crowdsourcer with the budget constraint $B = 16$ and the deadline $T = 8$. There are five users arriving online before the deadline with types $\theta_i = (a_i, d_i, \Gamma_i, c_i)$, where $a_i = d_i$, and $\Gamma_i$ can be omitted by assuming that each user has the same marginal value 1. Here the types $(a_i, d_i, c_i)$ of the five users are: $\theta_1 = (1, 1, 2)$, $\theta_2 = (2, 2, 4)$, $\theta_3 = (4, 4, 5)$, $\theta_4 = (6, 6, 1)$, and $\theta_5 = (7, 7, 3)$.*

We set $\epsilon = 1/2$ and $\delta = 1$. Then the *OMZ* mechanism works as follows.

◇ $t = 1$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (1, 2, \emptyset, 1/2, \emptyset)$, $V_1(\mathcal{S})/b_1 = 1/2$, thus $p_1 = 2$, $\mathcal{S} = \{1\}$, $\mathcal{S}' = \{1\}$. Update the density threshold: $\rho^* = 1/2$.

◇ $t = 2$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (2, 4, \{1\}, 1/2, \{1\})$, $V_2(\mathcal{S})/b_2 = 1/4$, thus $p_2 = 0$, $\mathcal{S}' = \{1, 2\}$. Update the density threshold: $\rho^* = 1/4$.

◇ $t = 4$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (4, 8, \{1, 2\}, 1/4, \{1\})$, $V_3(\mathcal{S})/b_3 = 1/5$, thus $p_3 = 0$, $\mathcal{S}' = \{1, 2, 3\}$. Update the density threshold: $\rho^* = 1/4$.

◇ $t = 6$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (8, 16, \{1, 2, 3\}, 1/4, \{1\})$, $V_4(\mathcal{S})/b_4 = 1$, thus $p_4 = 4$, $\mathcal{S} = \{1, 4\}$, $\mathcal{S}' = \{1, 2, 3, 4\}$.

◇ $t = 7$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (8, 16, \{1, 2, 3, 4\}, 1/4, \{1, 4\})$, $V_5(\mathcal{S})/b_5 = 1/3$, thus $p_5 = 4$. Finally, the set of selected users is $\mathcal{S} = \{1, 4, 5\}$, and the payments of these selected 3 users are 2, 4, 4 respectively.

### B. Mechanism Analysis

In the following, we will first prove that *OMZ* satisfies the computational efficiency (Lemma 2), individual rationality (Lemma 3), budget feasibility (Lemma 4), cost-truthfulness (Lemma 5), and consumer sovereignty (Lemma 6). Then, we will prove that *OMZ* can achieve a constant competitive ratio under both the i.i.d. model (Lemma 7) and the secretary model (Lemma 10) by elaborately fixing different values of $\delta$.

**Lemma 2.** *The OMZ mechanism is computationally efficient.*

*Proof:* Since the mechanism runs online, we only need to focus on the computation complexity at each time step $t \in \{1, \ldots, T\}$. Computing the marginal value of user $i$ takes $O(|\Gamma_i|)$ time, which is at most $O(m)$. Thus, the running time of computing the allocation and payment of user $i$ (lines 3-8) is bounded by $O(m)$. Next, we analyze the complexity of computing the density threshold (Algorithm 2). Finding the user with maximum marginal density takes $O(m|\mathcal{S}'|)$ time. Since there are $m$ tasks and each selected user should contribute at least one new task, the number of winners is at most $\min\{m, |\mathcal{S}'|\}$. Thus, the running time of Algorithm 2 is bounded by $O(m|\mathcal{S}'| \min\{m, |\mathcal{S}'|\})$. Therefore, the computation complexity at each time step (lines 3-13) is bounded by $O(m|\mathcal{S}'| \min\{m, |\mathcal{S}'|\})$. At the last stage, the sample set $\mathcal{S}'$ has the maximum number of samples, being $n/2$ with high probability. Thus, the computation complexity at each time step is bounded by $O(mn \min\{m, n\})$. ∎

Note that the above analysis is very conservative. In practice, the running time $O(|\Gamma_i|)$ of computing the marginal value is much less than $O(m)$. Moreover, the running time of *OMZ* will increase linearly with $n$ especially when $n$ is large.

**Lemma 3.** *The OMZ mechanism is individually rational.*

*Proof:* From the lines 4-6 of Algorithm 1, we know that $p_i \geq b_i$ if $i \in \mathcal{S}$, otherwise $p_i = 0$. Thus, we have $u_i \geq 0$. ∎

**Lemma 4.** *The OMZ mechanism is budget feasible.*

*Proof:* At each stage $i \in \{1, 2, \ldots, \lfloor \log_2 T \rfloor, \lfloor \log_2 T \rfloor + 1\}$, the mechanism uses a stage-budget $B' = \frac{2^{i-1}B}{2^{\lfloor \log_2 T \rfloor}}$. From the lines 4-5 of Algorithm 1, it is guaranteed that the current total payment does not exceed the stage-budget $B'$. Specially, the budget constraint of the last stage is $B$. Therefore, every stage is budget feasible, and when the deadline $T$ arrives, the total payment does not exceed $B$. ∎

Designing a cost-truthful mechanism relies on the rationale of *bid-independence*. Let $b_{-i}$ denote the sequence of bids arriving before the $i$-th bid $b_i$, i.e., $b_{-i} = (b_1, \ldots, b_{i-1})$. We call such a sequence *prefixal*. Let $p'$ be a function from prefixal sequences to prices (non-negative real numbers). We extend the definition of bid-independence [25] to the online scenario:

**Definition 2** (Bid-independent Online Auction). *An online auction is called bid-independent if the allocation and payment rules for each player $i$ satisfy:*

    *a) The auction constructs a price schedule $p'(b_{-i})$;*
    *b) If $p'(b_{-i}) \geq b_i$, player $i$ wins at price $p_i = p'(b_{-i})$;*
    *c) Otherwise, player $i$ is rejected, and $p_i = 0$.*

**Proposition 1.** *( [23], Proposition 2.1) An online auction is cost-truthful if and only if it is bid-independent.*

**Lemma 5.** *The OMZ mechanism is cost-truthful.*

*Proof:* Consider a user $i$ that arrives at some stage for which the density threshold is $\rho^*$. If by the time the user arrives there are no remaining budget, then the user's cost declaration will not affect the allocation of the mechanism and thus cannot improve his/her utility by submitting a false cost. Otherwise, assume there are remaining budget by the time the user arrives. In case $c_i \leq V_i(\mathcal{S})/\rho^*$, reporting any cost below $V_i(\mathcal{S})/\rho^*$ would not make a difference in the user's allocation and payment and his/her utility would be $V_i(\mathcal{S})/\rho^* - c_i \geq 0$. Declaring a cost above $V_i(\mathcal{S})/\rho^*$ would make the worker lose the auction, and his/her utility would be 0. In case $c_i > V_i(\mathcal{S})/\rho^*$, declaring any cost above $V_i(\mathcal{S})/\rho^*$ would leave the user unallocated with utility 0. If the user declares a cost lower than $V_i(\mathcal{S})/\rho^*$ he/she will be allocated. In such a case, however, his/her utility will be negative. Hence the user's utility is always maximized by reporting his/her true cost: $b_i = c_i$. ∎

**Lemma 6.** *The OMZ mechanism satisfies the consumer sovereignty.*

*Proof:* Each stage is an accepting process as well as a sampling process ready for the next stage. As a result, users are not automatically rejected during the sampling process, and are allocated as long as their marginal densities are not less than the current density threshold, and the allocated stage-budget has not been exhausted. ∎

Before analyzing the competitiveness of the *OMZ* mechanism, we first introduce an offline mechanism proposed by Singer [26], which has been proved to satisfy the *computational efficiency*, *individual rationality*, *budget feasibility*, and *truthfulness*. This mechanism does not have knowledge about users' costs, but it is an offline mechanism, i.e., all users submit their bids to the mechanism and wait for the mechanism to collect all the bids and decide on an allocation. This mechanism has been proved to be $O(1)$-*competitive* in maximizing the value of services received under budget constraint compared with the optimal solution. Therefore, we only need to prove that *OMZ* has a constant competitive ratio compared with this offline mechanism, then *OMZ* will also be $O(1)$-*competitive* compared with the optimal solution. Note that in the offline scenario satisfying the *time-truthfulness*

and *consumer sovereignty* is trivial, since all decisions are made after that all users' information is submitted to the crowdsourcer.

The offline mechanism adopts a *proportional share allocation rule*. As described in Algorithm 3, it consists of two phases: the *winner selection* phase and the *payment determination* phase. The *winner selection* phase has the same working process as Algorithm 2. To compute the payment for each winner $i \in \mathcal{S}$, we sort the users in $\mathcal{U} \setminus \{i\}$:

$$\frac{V_{i_1}(\mathcal{Q}_0)}{b_{i_1}} \geq \frac{V_{i_2}(\mathcal{Q}_1)}{b_{i_2}} \geq \cdots \geq \frac{V_{i_{n-1}}(\mathcal{Q}_{n-2})}{b_{i_{n-1}}},$$

where $V_{i_j}(\mathcal{Q}_{j-1}) = V(\mathcal{Q}_{j-1} \cup \{i_j\}) - V(\mathcal{Q}_{j-1})$ denotes the marginal value of the $j$-th user and $\mathcal{Q}_j$ denotes the first $j$ users according to this sorting over $\mathcal{U} \setminus \{i\}$ and $\mathcal{Q}_0 = \emptyset$. The marginal value of user $i$ at position $j$ is $V_{i(j)}(\mathcal{Q}_{j-1}) = V(\mathcal{Q}_{j-1} \cup \{i\}) - V(\mathcal{Q}_{j-1})$. Let $k'$ denote the position of the last user $i_j \in \mathcal{U} \setminus \{i\}$, such that $b_{i_j} \leq V_{i_j}(\mathcal{Q}_{j-1})B/V(\mathcal{Q}_j)$. For brevity we will write $b_{i(j)} = V_{i(j)}(\mathcal{Q}_{j-1})b_{i_j}/V_{i_j}(\mathcal{Q}_{j-1})$, and $\eta_{i(j)} = V_{i(j)}(\mathcal{Q}_{j-1})B/V(\mathcal{Q}_{j-1} \cup \{i\})$. In order to guarantee the truthfulness, each winner should be paid the critical value, which means that user $i$ would not win the auction if he/she bids higher than this value. Thus, the payment for user $i$ should be the maximum of these $k' + 1$ prices:

$$p_i = \max_{j \in [k'+1]} \{\min\{b_{i(j)}, \eta_{i(j)}\}\}.$$

---

**Algorithm 3:** Proportional Share Mechanism (Offline) [26]

**Input**: Budget constraint $B$, User set $\mathcal{U}$
    /* Winner selection phase             */
1  $\mathcal{S} \leftarrow \emptyset$; $i \leftarrow \arg\max_{j \in \mathcal{U}}(V_j(\mathcal{S})/b_j)$;
2  **while** $b_i \leq \frac{V_i(\mathcal{S})B}{V(\mathcal{S} \cup \{i\})}$ **do**
3      $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$;
4      $i \leftarrow \arg\max_{j \in \mathcal{U} \setminus \mathcal{S}}(V_j(\mathcal{S})/b_j)$;
5  **end**
    /* Payment determination phase       */
6  **foreach** $i \in \mathcal{U}$ **do** $p_i \leftarrow 0$;
7  **foreach** $i \in \mathcal{S}$ **do**
8      $\mathcal{U}' \leftarrow \mathcal{U} \setminus \{i\}$; $\mathcal{Q} \leftarrow \emptyset$;
9      **repeat**
10         $i_j \leftarrow \arg\max_{j \in \mathcal{U}' \setminus \mathcal{Q}}(V_j(\mathcal{Q})/b_j)$;
11         $p_i \leftarrow \max\{p_i, \min\{b_{i(j)}, \eta_{i(j)}\}\}$;
12         $\mathcal{Q} \leftarrow \mathcal{Q} \cup \{i_j\}$;
13      **until** $b_{i_j} \leq \frac{V_{i_j}(\mathcal{Q}_{j-1})B}{V(\mathcal{Q})}$;
14  **end**
15  **return** $(\mathcal{S}, p)$;

---

Let $Z$ be the set of selected users $\mathcal{S}$ computed by Algorithm 3, and the value of $Z$ is $V(Z)$. The density of $Z$ is $\rho = V(Z)/B$. Define $Z_1$ and $Z_2$ as the subsets of $Z$ that appears in the first and second half of the input stream, respectively. When the stage $\lfloor \log_2 T \rfloor$ is over, we obtain the sample set $\mathcal{S}'$ consisting of all users arriving before the time $\lfloor T/2 \rfloor$. Thus, we have $Z_1 = Z \cap \mathcal{S}'$, and $Z_2 = Z \cap \{\mathcal{U} \setminus \mathcal{S}'\}$. Let $Z_1'$ denote the set of selected users computed by Algorithm 2 [‡] based on the sample set $\mathcal{S}'$ and the allocated stage-budget $B/2$, and the

value of $Z_1'$ is $V(Z_1')$. The density of $Z_1'$ is $\rho_1' = 2V(Z_1')/B$. The density threshold of the last stage is $\rho^* = \rho_1'/\delta$. Let $Z_2'$ denote the set of selected users computed by Algorithm 1 at the last stage. Assume that the value of each user is at most $V(Z)/\omega$, where the parameter $\omega$ will be fixed later.

*1) Competitiveness Analysis under the I.I.D. Model:* Since the costs and values of all users are i.i.d., they can be selected in the set $Z$ with the same probability. The sample set $\mathcal{S}'$ is a random subset of $\mathcal{U}$ since all users arrive in a random order. Therefore the number of users from $Z$ in the sample set $\mathcal{S}'$ follows a hypergeometric distribution $H(n/2, |Z|, n)$. Thus, we have $\mathbb{E}[|Z_1|] = \mathbb{E}[|Z_2|] = |Z|/2$. The value of each user can be seen as an i.i.d. random variable, and because of the submodularity of $V(\mathcal{S})$, it can be derived that: $\mathbb{E}[V(Z_1)] = \mathbb{E}[V(Z_2)] \geq V(Z)/2$. The expected total payments to the users from both $Z_1$ and $Z_2$ are $B/2$. Since $V(Z_1')$ is computed with the stage-budget $B/2$, it can be derived that: $\mathbb{E}[V(Z_1')] \geq \mathbb{E}[V(Z_1)] \geq V(Z)/2$, and $\mathbb{E}[\rho_1'] \geq \rho$, where the first inequality follows from the fact that $V(Z_1')$ is the optimal solution computed by Algorithm 2 with stage-budget $B/2$ according to the *proportional share allocation* rule. Therefore, we only need to prove that the ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ is at least a constant, then the *OMZ* mechanism will also have a constant expected competitive ratio compared with the offline mechanism.

**Lemma 7.** *For sufficiently large $\omega$, the ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ is at least a constant. Specially, this ratio approaches $1/4$ as $\omega \to \infty$ and $\delta \to 4$.*

The proof of Lemma 7 is given in Appendix B.

*2) Competitiveness Analysis under the Secretary Model:*

**Lemma 8.** *( [24], Lemma 16) For sufficiently large $\omega$, the random variable $|V(Z_1) - V(Z_2)|$ is bounded by $V(Z)/2$ with a constant probability.*

Note that a non-negative submodular function is also a subadditive function, so we have $V(Z_1) + V(Z_2) \geq V(Z)$. Thus, Lemma 8 can be easily extended to the following corollary.

**Corollary 1.** *For sufficiently large $\omega$, both $V(Z_1)$ and $V(Z_2)$ are at least $V(Z)/4$ with a constant probability.*

**Lemma 9.** *Given a sample set $\mathcal{S}'$, the total value of selected users computed by Algorithm 2 with the budget $B'/2$ is at least a half of that computed with the budget $B'$.*

The proof of Lemma 9 is given in Appendix C.

Note that the total value of selected users from the sample set $\mathcal{S}'$ computed by Algorithm 2 with the budget $B$ is not less than $V(Z_1)$. Thus, considering Corollary 1 and Lemma 9, it can be derived that: $V(Z_1') \geq V(Z_1)/2 \geq V(Z)/8$. Therefore, it only needs to prove that the ratio of $V(Z_2')$ to $V(Z_1')$ is at least a constant, then the *OMZ* mechanism will also have a constant competitive ratio compared with the offline mechanism.

---

‡When we mention Algorithm 2 in the following analysis, it means the line 10 in Algorithm 1, which calls the GetDensityThreshold algorithm.

**Lemma 10.** *For sufficiently large $\omega$, the ratio of $V(Z_2')$ to $V(Z_1')$ is at least a constant. Specially, this ratio approaches $1/12$ as $\omega \to \infty$ and $\delta \to 12$.*

The proof of Lemma 10 is given in Appendix D.

From the above analysis, we know that the $OMZ$ mechanism has a competitive factor of at least 8 (96) of the offline *proportional share* solution under the i.i.d. model (the secretary model). While the competitive ratio may seem large, we emphasize that our goal is to show that the $OMZ$ mechanism is indeed $O(1)$-*competitive*, and thus its performance guarantee is independent of the parameters of the problem (e.g. number of users, their costs, the tasks they can complete, etc.). We will later show that the mechanism performs well in practice (see Section V), implying that bounded competitive ratio serves as a good guide for designing such mechanisms.

**Theorem 1.** *The OMZ mechanism satisfies computational efficiency, individual rationality, budget feasibility, truthfulness, consumer sovereignty, and constant competitiveness under the zero arrival-departure interval case.*

## IV. ONLINE MECHANISM UNDER GENERAL CASE

In this section, we consider the general case where each user may have a *non-zero arrival-departure interval*, and there may be multiple online users in the auction simultaneously. First, we change the settings of Example 1 to show that the *OMZ* mechanism is *not* time-truthful under the general case.

**Example 2.** *All the settings are the same as Example 1 except for that user 1 has a non-zero arrival-departure interval, $a_1 < d_1$. Specially, the type of user 1 is $\theta_1 = (1, 5, 2)$.*

In this example, if user 1 report his/her type truthfully, then he/she will obtain the payment 2 according to the *OMZ* mechanism. However, if user 1 delays announcing his/her arrival time and reports $\theta_1' = (5, 5, 2)$, then he/she will improve his/her payment to 8 according to the *OMZ* mechanism (the detailed computing process is omitted).

In the following, we will present a new online mechanism, *OMG*, and prove that it satisfies all six desirable properties under the general case.

### A. Mechanism Design

In order to hold desirable properties of *OMZ*, we adopt a similar algorithm framework under the general case. Meanwhile, in order to guarantee the *cost-* and *time-truthfulness*, it is necessary to modify *OMZ* based on three principles. First, any user is added to the sample set only when he/she departs; otherwise, the bid-independence will be destroyed if his/her arrival-departure time spans multiple stages, because a user can indirectly affect his/her payment now. Second, if there are multiple users who have not yet departed at some time, we sort these online users according to their marginal values, instead of marginal densities, and preferentially select those users with higher marginal value. In this way, the bid-independence can be held. Note that, this principle can be used to enable *OMZ* to adapt to the case when two users have the same arrival time. Third, whenever a new time step arrives,

we scan through the list of users who have not yet departed and selects those whose marginal densities are not less than the current density threshold under the stage-budget constraint, even if some arrived much earlier. At the departure time of any user who was selected as a winner, the user is paid for a price equal to the maximum price attained during the user's reported arrival-departure interval, even if this price is larger than the price at the time step when the user was selected as a winner.

According to the above principles, we design *OMG* satisfying all desirable properties under the general case, as described in Algorithm 4. Specially, we consider two cases. The first case is when the current time step $t$ is not at the end of any stage. In this case, the density threshold remains unchanged. The following operations (the lines 3-11 in Algorithm 4) are performed. First, all new users arriving at time step $t$ are added to a set of online users $\mathcal{O}$. Then we make decision on whether to select these online users one by one in the order of their marginal values; the users with higher marginal values will be selected first. If an online user $i$ has been selected as a winner before time step $t$, we need not to make decision on him/her again because he/she is impossible to obtain a higher payment than before (to be proved later in Lemma 13). Otherwise, we need to make decision on him/her again: if his/her marginal density is not less than the current density threshold, and the allocated stage-budget has not been exhausted, he/she will be selected as a winner. Meanwhile, we give user $i$ a payment $p_i = V_i(\mathcal{S})/\rho^*$, and add he/she to the set of selected users $\mathcal{S}$. Finally, we remove all users departing at time step $t$ from $\mathcal{O}$, and add them to the sample set $\mathcal{S}'$.

The second case is when the current time step is just at the end of some stage. In this case, the density threshold will be updated. The mechanism works as the lines 13-22. We need to make decision on whether to select these online users, and at what prices, one by one in the order of their marginal values, no matter whether they have ever been selected as the winners before time step $t$. As shown in the lines 17-20, if user $i$ can obtain a higher payment than before, his/her payment will be updated. Meanwhile, if user $i$ has never been selected as a winner before time step $t$, he/she will be added to the set $\mathcal{S}$.

Return to Example 2. If all of the five users report their types truthfully, then the *OMG* mechanism works as follows.

◇ $t = 1$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (1, 2, \emptyset, 1/2, \emptyset)$, $V_1(\mathcal{S})/b_1 = 1/2$, thus $p_1 = 2$, $\mathcal{S} = \{1\}$. Update the density threshold: $\rho^* = 1/2$, $p_1$ remains unchanged.

◇ $t = 2$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (2, 4, \emptyset, 1/2, \{1\})$, $V_2(\mathcal{S})/b_2 = 1/4$, thus $p_2 = 0$, $\mathcal{S}' = \{2\}$. Update the density threshold: $\rho^* = 1/4$, increase $p_1$ to 4.

◇ $t = 4$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (4, 8, \{2\}, 1/4, \{1\})$, $V_3(\mathcal{S})/b_3 = 1/5$, thus $p_3 = 0$, $\mathcal{S}' = \{2, 3\}$. Update the density threshold: $\rho^* = 1/8$, increase $p_1$ to 8.

◇ $t = 5$: user 1 departs, so $\mathcal{S}' = \{1, 2, 3\}$.

◇ $t = 6$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (8, 16, \{1, 2, 3\}, 1/8, \{1\})$, $V_4(\mathcal{S})/b_4 = 1$, thus $p_4 = 8$, $\mathcal{S} = \{1, 4\}$, $\mathcal{S}' = \{1, 2, 3, 4\}$.

◇ $t = 7$: $(T', B', \mathcal{S}', \rho^*, \mathcal{S}) = (8, 16, \{1, 2, 3, 4\}, 1/8, \{1, 4\})$, $V_5(\mathcal{S})/b_5 = 1/3$, thus $p_5 = 0$, $\mathcal{S}' = \{1, 2, 3, 4, 5\}$.

Thus, user 1 can obtain the payment 8 according to the *OMG* mechanism. Even if user 1 delays announcing his/her arrival time and reports $\theta'_1 = (5, 5, 2)$, he/she still cannot

---

**Algorithm 4:** Online Mechanism under General Case (*OMG*)

**Input**: Budget constraint $B$, deadline $T$

1   $(t, T', B', \mathcal{S}', \rho^*, \mathcal{S}) \leftarrow (1, \frac{T}{2^{\lfloor \log_2 T \rfloor}}, \frac{B}{2^{\lfloor \log_2 T \rfloor}}, \emptyset, \epsilon, \emptyset)$;

2   **while** $t \leq T$ **do**

3     Add all new users arriving at time step $t$ to a set of online users $\mathcal{O}$; $\mathcal{O}' \leftarrow \mathcal{O} \setminus S$;

4     **repeat**

5       $i \leftarrow \arg\max_{j \in \mathcal{O}'} (V_j(\mathcal{S}))$;

6       **if** $b_i \leq V_i(\mathcal{S})/\rho^* \leq B' - \sum_{j \in \mathcal{S}} p_j$ **then**

7         $p_i \leftarrow V_i(\mathcal{S})/\rho^*$; $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$;

8       **else** $p_i \leftarrow 0$;

9       $\mathcal{O}' \leftarrow \mathcal{O}' \setminus \{i\}$;

10    **until** $\mathcal{O}' = \emptyset$;

11    Remove all users departing at time step $t$ from $\mathcal{O}$, and add them to $\mathcal{S}'$;

12    **if** $t = \lfloor T' \rfloor$ **then**

13      $\rho^* \leftarrow$ **GetDensityThreshold**$(B', \mathcal{S}')$;

14      $T' \leftarrow 2T'$; $B' \leftarrow 2B'$; $\mathcal{O}' \leftarrow \mathcal{O}$;

15      **repeat**

16        $i \leftarrow \arg\max_{j \in \mathcal{O}'} (V_j(\mathcal{S} \setminus \{j\}))$;

17        **if** $b_i \leq V_i(\mathcal{S} \setminus \{i\})/\rho^* \leq B' - \sum_{j \in \mathcal{S}} p_j + p_i$ **and** $V_i(\mathcal{S} \setminus \{i\})/\rho^* > p_i$ **then**

18          $p_i \leftarrow V_i(\mathcal{S} \setminus \{i\})/\rho^*$;

19          **if** $i \notin \mathcal{S}$ **then** $\mathcal{S} \leftarrow \mathcal{S} \cup \{i\}$;

20        **end**

21        $\mathcal{O}' \leftarrow \mathcal{O}' \setminus \{i\}$;

22      **until** $\mathcal{O}' = \emptyset$;

23    **end**

24    $t \leftarrow t + 1$;

25 **end**

---

improve his/her payment (the detailed computing process is omitted). Therefore, the time-truthfulness can be guaranteed in this case.

### B. Mechanism Analysis

It is convenient to prove that the *OMG* mechanism also holds the *individual rationality*, *consumer sovereignty*, and *constant competitiveness* as *OMZ* (with almost the same proof), although *OMG* may have slightly lower competitive ratio than *OMZ*. In the following, we prove that *OMG* also satisfies the *computational efficiency*, the *budget feasibility*, and most importantly, the *cost- and time-truthfulness*.

**Lemma 11.** *The OMG mechanism is computationally efficient.*

*Proof:* The *OMG* mechanism needs to compute the allocations and payments of multiple online users at each time step. Thus, the running time of computing the allocations and payments at each time step is bounded by $O(m|\mathcal{O}|) < O(mn)$, where $|\mathcal{O}|$ is the number of online users. The complexity of computing the density threshold is the same as that of *OMZ*. Thus, the computation complexity at each time step is the same as that of *OMZ*, i.e., bounded by $O(mn \min\{m, n\})$. ∎

**Lemma 12.** *The OMG mechanism is budget feasible.*

*Proof:* From the lines 6-7 and 17-18 of Algorithm 4, it is guaranteed that the current total payment does not exceed the stage-budget $B'$. Note that in the line 17, $p_i$ is the price paid for user $i$ in the previous stage instead of the current stage,

so it cannot lead to the overrun of the current stage-budget. Thus every stage is budget feasible, and when the deadline $T$ arrives, the total payment does not exceed $B$. ∎

**Lemma 13.** *The OMG mechanism is cost- and time-truthful.*

The proof of Lemma 13 is given in Appendix E.

**Theorem 2.** *The OMG mechanism satisfies the computational efficiency, individual rationality, budget feasibility, truthfulness, consumer sovereignty, and constant competitiveness under the general case.*

## V. PERFORMANCE EVALUATION

To evaluate the performance of our online mechanisms, we implemented *OMZ* and *OMG*, and compared them against the following three benchmarks. The first benchmark is the (approximate) *optimal* offline solution which has full knowledge about all users' types. The problem in this scenario is essentially a *budgeted maximum coverage problem*, which is a well-known NP-hard problem. It is known that a greedy algorithm provides a $(1 - 1/e)$-approximation solution [27]. The second benchmark is the *proportional share* mechanism in the offline scenario (Algorithm 3). The third benchmark is the *random* mechanism, which adopts a naive strategy, i.e., rewards users based on an uninformed fixed density threshold. The performance metrics include the *running time*, the *crowdsourcer's value* and the *truthfulness*.
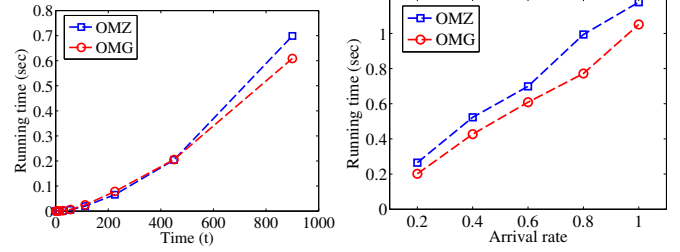
### A. Simulation Setup

Simulation Setting 1: First, we consider a Wi-Fi signal sensing application with the same scenario as [28]. As shown in Fig. 3 obtained from the Google Map, the RoI is located at Manhattan, NY, including three avenues of 0.319km length and three streets of 1.135km length. We divide each road in the RoI into discrete PoIs with a uniform spacing of 1m, so the RoI consists of 4353 PoIs ($m = 4352$) in total. Without loss of generality, let the coverage requirement of each PoI be 1. We set the deadline ($T$) to 1800s, and vary the budget ($B$) from 100 to 10000 with the increment of 100. Users arrive according to a Poisson process in time with arrival rate $\lambda$. We vary $\lambda$ from 0.2 to 1 with the increment of 0.2. Whenever a user arrives, he/she is placed at a random location on the roads. In *OMZ* each user has zero arrival-departure interval, and in *OMG* the arrival-departure interval of each user is uniformly distributed over $[0, 300]$ seconds. The sensing range ($R$) of each sensor is set to 7 meters. Each user's cost is uniformly distributed over $[1, 10]$. The initial density threshold ($\epsilon$) of Algorithm 1 and 4 is set to 1 §. As we proved in Lemma 7, when $\delta = 4$ *OMZ* is $O(1)$-competitive for sufficiently large $\omega$. Meanwhile we note that $\omega$ increases with the number of users who have arrived. Thus, we set $\delta = 1$ initially, and change it to $\delta = 4$ once the size of the sample set exceeds a specified threshold. Note that this threshold could be an empirical value, which is decided by when each user's value is a very small fraction of the total value of all users in the sample set. In our simulation, we

§ $\epsilon$ only affects a very small number of users arriving at the first stage, and thus its impacts on the results can be omitted.



Fig. 3. The region of interest.



(a) At different stages ($\lambda = 0.6$)  (b) Impact of $\lambda$ (at the last phase)

Fig. 4. Running Time.

set this threshold to 240, because we observe that each user's value is at most $1/100$ of the total value when the number of users is larger than 240. For the *random* mechanism, we obtain the average performance of 50 such solutions for evaluations, where in each solution the density threshold was chosen at random from the range of 1 to 29 ¶.

Simulation Setting 2: Second, we consider a simulation scenario with real human mobility traces collected from KAIST [29]. Altogether 92 daily trajectories with GPS positions are collected by participants (users) in a region of $8000 \times 14000\text{m}^2$. We divide the region into discrete PoIs with a uniform spacing of 100m. We set the deadline ($T$) to 1800s, and vary the budget ($B$) from 50 to 1000 with the increment of 50. Users arrive according to a Poisson process in time with arrival rate $\lambda = 0.05$. Whenever a user arrives, he/she submits a set of tasks (PoIs) according to his/her mobility traces in the future 10 minutes. The sensing range ($R$) of each sensor (user) is set to 50m. We fix the value of $\delta$ to 4 as the lower-bound. Other parameters are the same to Simulation Setting 1.

Note that all the evaluation results in this section are based on Simulation Setting 1 unless otherwise specified, and Simulation Setting 2 is only used to evaluate crowdsourcer's value. All the simulations were run on a PC with 1.7 GHz CPU and 8 GB memory. Each measurement is averaged over 100 instances.

### B. Evaluation Results

Running Time: Fig. 4 shows the running time of *OMZ* and *OMG*. Specially, Fig. 4(a) plots the running time at different stages while $\lambda = 0.6$‖. Fig. 4(b) plots the running time at the last stage with different arrival rates ($\lambda$). Both the *OMZ* and *OMG* mechanisms have similar performance while *OMG* outperforms *OMZ* slightly. Note that the size of the sample set increases linearly with the time $t$ and the arrival rate $\lambda$, so

¶ Each user can cover at most 29 PoIs, and his/her bid is at least 1, so his/her marginal density is at most 29.

‖ As we proved in Lemma 2, the computation complexity is dominated by computing the density threshold, so only the running time at the end time of each stage is plotted.

(a) Impact of $B$ ($\lambda = 0.6$)  (b) Impact of $\lambda$ ($B = 2000$)



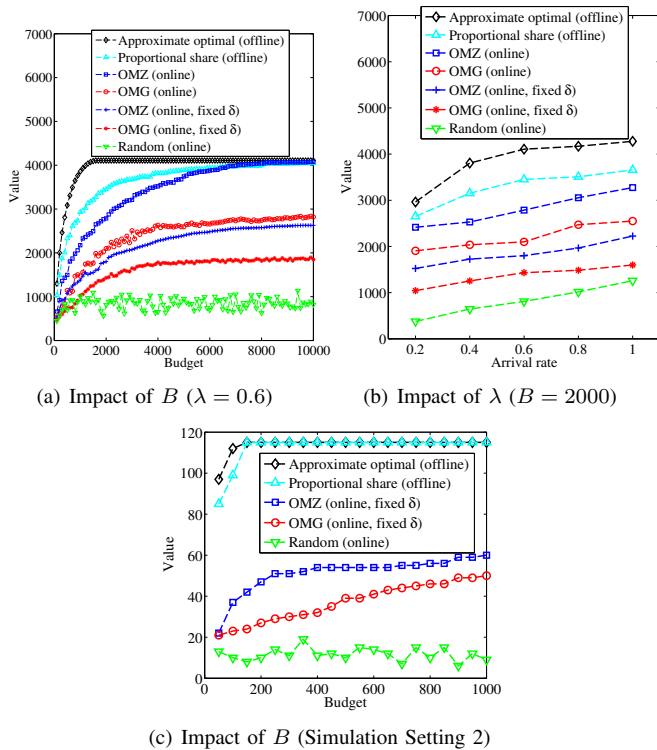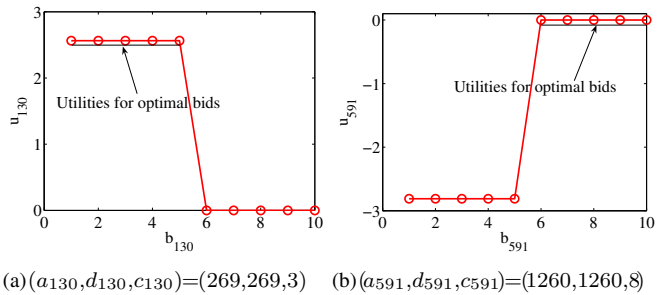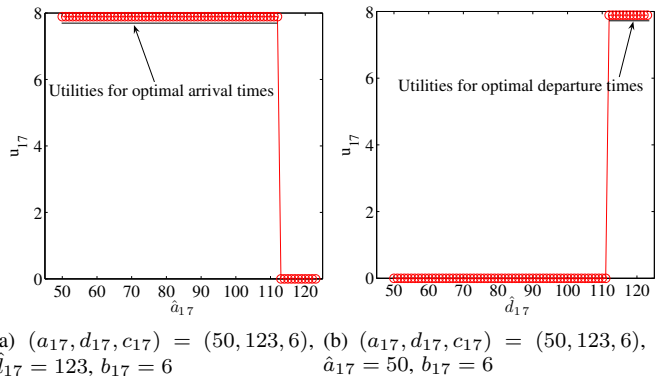(c) Impact of $B$ (Simulation Setting 2)

Fig. 5.  Crowdsourcer's value.

Fig. 4 implies the relationship between the running time and the number of users. Thus, from Fig. 4 we can infer that the running time increases linearly with the number of users ($n$), which is consistent with our analysis in Section III-B.

Crowdsourcer's Value: Fig. 5 compares the crowdsourcer's value achieved by *OMZ* and *OMG* against the three benchmarks. In order to investigate the impact of $\delta$ on *OMZ* and *OMG*, we also provided the simulation results by fixing the value of $\delta$ to 4 as the lower-bound. From Fig. 5(a) and Fig. 5(c) we can observe that the crowdsourcer obtains higher value when the budget constraint increases. From Fig. 5(b) we can observe that the crowdsourcer obtains higher value when more users participate. The *approximate optimal* mechanism and the *proportional share* mechanism operate in the offline scenario, where the true types or strategies of all users are known a priori, and will therefore always outperform *OMZ* and *OMG*. It is shown that the *proportional share* mechanism sacrifices some value of the crowdsourcer to achieve the cost-truthfulness compared with the *approximate optimal* mechanism, and *OMG* also sacrifices some value to achieve the time-truthfulness compared with *OMZ*. We can also observe that both *OMZ* and *OMG* are guaranteed to be within a constant factor of the offline solutions, although the crowdsourcer's value will be reduced to some extent by fixing the value of $\delta$ without any heuristic method. Specially, although both *OMZ* and *OMG* are only guaranteed to be within a competitive factor of at least 8 of the *proportional share* solution in expectation as we proved in Lemma 7, the simulation results show that this ratio is almost as small as 1.6 for *OMZ* or 2.4 for *OMG* by setting the value of $\delta$ heuristically. As compared to the *approximate optimal* solution, this ratio is still below 2.2 for *OMZ* or below 3.4 for *OMG*. In addition, we can see that both



(a) $(a_{130}, d_{130}, c_{130}) = (269, 269, 3)$  (b) $(a_{591}, d_{591}, c_{591}) = (1260, 1260, 8)$

Fig. 6.  Cost-truthfulness of *OMZ*.



(a) $(a_{17}, d_{17}, c_{17}) = (50, 123, 6)$, (b) $(a_{17}, d_{17}, c_{17}) = (50, 123, 6)$, $\hat{d}_{17} = 123$, $b_{17} = 6$  $\hat{a}_{17} = 50$, $b_{17} = 6$



(c) $(a_{85}, d_{85}, c_{85}) = (201, 476, 4)$, $\hat{d}_{85} = 476$, $b_{85} = 4$

Fig. 7.  Time-truthfulness of *OMG*.

*OMZ* and *OMG* largely outweigh the *random* mechanism.

Truthfulness: We first verified the cost-truthfulness of *OMZ* by randomly picking two users (ID=130 and ID=591) and allowing them to bid prices that are different from their true costs. We illustrate the results in Fig. 6. As we can see, user 130 achieves his/her optimal utility if he/she bids truthfully ($b_{130} = c_{130} = 3$) in Fig. 6(a) and user 591 achieves his/her optimal utility if he/she bids truthfully ($b_{591} = c_{591} = 8$) in Fig. 6(b). Then we further verified the time-truthfulness of *OMG* by randomly picking two users (ID=17 and ID=85) and allowing them to report their arrival/departure times that are different from their true arrival/departure times. We illustrate the results in Fig. 7. As shown in Fig. 7(a) and Fig. 7(b), user 17 achieves his/her optimal utility if he/she reports his/her true arrival and departure times ($\hat{a}_{17} = a_{17} = 50$, $\hat{d}_{17} = d_{123} = 50$). As shown in Fig. 7(c), user 85 achieves his/her optimal utility if he/she reports his/her true arrival time ($\hat{a}_{85} = a_{85} = 201$). Note that reporting any departure time ($a_{85} \leq \hat{d}_{85} \leq d_{85}$) does not affect the utility of user 85.

## VI. Discussion on Frugal and Profit-Maximizing Online Incentive Mechanisms

Although this work mainly focuses on the *budget feasible* mechanisms, we would like to emphasize that our main idea and framework can also be applied to design online incentive mechanisms with other objectives such as frugality and profit maximization, while guaranteeing both consumer sovereignty and time-truthfulness that are seldom considered by existing work.

First, we consider the objective of *frugality*, namely that the crowdsourcer expects to minimize the total payment while obtaining the specific value of services $\Psi$. Based on our system model, it can be expressed as follows.

$$\textbf{Minimize } \sum_{i \in \mathcal{S}} p_i \textbf{ subject to } V(\mathcal{S}) \geq \Psi.$$

Similarly, the frugal online incentive mechanisms should satisfy the *computational efficiency*, *individual rationality*, *truthfulness*, *consumer sovereignty*, and two other properties that are different from budget feasible mechanisms, namely the *task completeness* and *constant frugality*. The *task completeness* means that the crowdsourcer can obtain the specific value of services, and the *constant frugality* means that the total payment made by the crowdsourcer has a constant approximate ratio compared to the minimal cost required for obtaining the specific value in the offline scenario. In order to guarantee the *consumer sovereignty*, we can adopt a multiple-stage sampling-accepting process similar to the budget feasible mechanisms. The difference is that each stage $i$ is allocated a stage-task $\Psi' = 2^{i-1}\Psi/2^{\lfloor \log_2 T \rfloor}$, meaning that at each stage the crowdsourcer should just obtain the value $\Psi'$. In order to satisfy *task completeness* and *constant frugality*, it is intuitive that we can dynamically learn a budget that is enough for allocating users to obtain a specific value of services, then use this budget to compute a threshold by leveraging budget feasible mechanisms, and finally use this threshold for making further decisions. We have used this idea to design constant-frugal mechanisms with a linear value function in our latest work [30]. This idea is also applicable to frugal mechanisms with a submodular value function, but it is still challenging to obtain a satisfactory frugality ratio.

Second, another research line is to design online mechanisms for maximizing the *profit* of the crowdsourcer, namely the value of services minus the total payment, without budget constraint, where the profit function is non-monotone submodular, and can be negative. Followed by this work, another latest work [31] has designed constant-competitive profit-maximizing online incentive mechanisms to satisfy both *consumer sovereignty* and *time-truthfulness* by leveraging similar ideas.

## VII. Related Work

### A. Mechanism Design for Mobile Crowd Sensing

Reddy et al. [32] developed recruitment frameworks to enable the crowdsourcer to identify well-suited participants for data collections. However, they focused only on the user selection instead of the incentive mechanism design. At present, there are many studies [10]–[16] on incentive mechanism design for MCS applications in the offline scenario. Generally, two system models are considered: the platform/crowdsourcer-centric model where the crowdsourcer provides a fixed reward to participating users, and the user-centric model where users can have their expected prices for the sensing service. For the crowdsourcer-centric model, incentive mechanisms were designed by using a Stackelberg game [12], [13]. The Nash Equilibrium and Stackelberg Equilibrium were computed as the solution, where the costs of all users or their probability distribution was assumed to be known. In contrast, the user-centric model allows that each user has a private cost only known to itself. Danezis et al. [10] developed a sealed-bid second-price auction to estimate the users' value of sensing data with location privacy. Lee and Hoh [11] designed and evaluated a reverse auction based dynamic price incentive mechanism, where users can sell their sensed data to a service provider with users' claimed bids. Jaimes et al. [14] proposed a recurrent reverse auction incentive mechanism with a greedy algorithm that selects a representative subset of the users according to their location given a fixed budget. Yang et al. [13] designed an auction-based incentive mechanism, and proved this mechanism was computationally efficient, individually rational, profitable, and truthful. Feng et al. [15] investigated a more complex auction-based incentive mechanism by considering the crucial dimension of location information when assigning sensing tasks to smartphones. Luo et al. [16] designed a profit-maximizing incentive mechanism based on all-pay auctions. However, all of these studies failed to account for the online arrival of users.

More recently, there are some studies on online mechanism design for crowdsourcing markets [17]–[19]. Singer et al. [17] and Singla et al. [18] presented pricing mechanisms for crowdsourcing markets based on the bidding model and the posted price model respectively. However, they focused only on a simple additive utility function instead of the submodular one. Only Badanidiyuru et al. [19] considered pricing mechanisms for maximizing the submodular utility function. But they mainly focused on the posted price mechanisms, and only a small part of content discusses the bidding mechanisms under the secretary model but not the i.i.d. model. Moreover, they failed to consider the *consumer sovereignty*. In addition, none of the above studies considered the *time-truthfulness*.

### B. Online Auctions and Generalized Secretary Problems

*Online auction* is the essence of many networked markets, in which information about goods, agents, and outcomes is revealed one by one online in a random order, and the agents must make irrevocable decisions without knowing future information. Combining optimal stopping theory with game theory provides us a powerful tool to model the actions of rational agents applying competing stopping rules in an online auction. The theory of *optimal stopping* is concerned with the problem of choosing a time to take a particular action, in order to maximize an expected reward or minimize an expected cost. A classic problem of optimal stopping theory is the *secretary problem*: designing an algorithm for hiring

one secretary from a pool of $n$ applicants arriving online, to maximize the probability of hiring the best secretary. Many variants of the classic secretary problem have been studied in the literature, and the most relevant to this work is the *knapsack secretary problem*, in which each applicant also has a cost and the goal is to maximize performance of the secretarial group as along as the total cost of selected applicants does not exceed a given budget. Babaioff et al. [33] and Bateni et al. [24] respectively presented constant competitive algorithms for the linear knapsack secretary problem in which the objective function is linear, and the *submodular knapsack secretary problem* in which the objective function is submodular. Our problem is similar to the *submodular knapsack secretary problem* in form, but we need to consider two significant properties, the *truthfulness* and the *consumer sovereignty*. Although some solutions ( [21], [22], [34]) of online auctions provided good ideas of designing truthful mechanisms, they cannot be directly applied to the problem setting with submodular value function and budget constraint. Moreover, none of these solutions considered the *consumer sovereignty*.

## VIII. CONCLUSIONS

In this paper, we have designed online incentive mechanisms to motivate mobile users to participate in mobile crowd sensing, which is a new sensing paradigm allowing us to efficiently collect data for numerous novel applications. Compared with existing offline incentive mechanisms, we focus on a more real scenario where users arrive one by one online. We have modeled the problem as an online auction in which the users submit their private types to the crowdsourcer over time, and the crowdsourcer aims at selecting a subset of users before a specified deadline for maximizing the total value of services provided by selected users under a budget constraint. We focus on the monotone submodular value function that can be applied in many realistic scenarios. Two online mechanisms have been designed under different assumptions: *OMZ* can be applied to the zero arrival-departure interval case where the arrival time of each user equals to his/her departure time, and *OMG* can be applied to a more general case. We have proven that our mechanisms satisfy the *computational efficiency*, *individual rationality*, *budget feasibility*, *truthfulness*, *consumer sovereignty* and *constant competitiveness*. Besides, our main idea and framework can also be applied to design online incentive mechanisms with other objectives such as frugality and profit maximization.

## APPENDIX

### A. Proof of Lemma 1

Considering $V(\mathcal{S}) = \sum_{j=1}^{m} \min\{r_j, \sum_{i\in\mathcal{S}} v_{i,j}\}$, for any $X \subseteq Y \subseteq \mathcal{U}$ and $x \in \mathcal{U}\backslash Y$ we have

$$V(X \cup \{x\}) - V(X) = \sum_{j=1}^{m} \min\{\max\{0, r_j - \sum_{i\in X} v_{i,j}\}, v_{x,j}\}$$
$$\geq \sum_{j=1}^{m} \min\{\max\{0, r_j - \sum_{i\in Y} v_{i,j}\}, v_{x,j}\}$$
$$= V(Y \cup \{x\}) - V(Y).$$



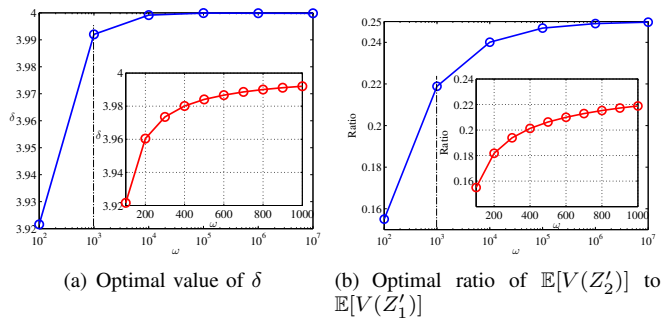(a) Optimal value of $\delta$    (b) Optimal ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$

Fig. 8. The optimal ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ by fixing proper $\delta$ with different values of $\omega$.

Moreover, for any $X \subseteq \mathcal{U}$ and $x \in \mathcal{U}\backslash X$ we have $V(X \cup \{x\}) - V(X) \geq 0$. Therefore $V(\mathcal{S})$ is monotone submodular by Definition 1.

### B. Proof of Lemma 7

We consider two cases according to the total payment to the selected users at the last stage as follows.

**Case (a):** The total payment to the selected users at the last stage is at least $\alpha B$, $\alpha \in (0, 1/2]$. In this case, since each selected user has marginal density at least $\rho^*$, so we have that

$$V(Z_2') \geq \rho^*\alpha B = \frac{\alpha\rho_1' B}{\delta} = \frac{2\alpha V(Z_1')}{\delta}.$$

**Case (b):** The total payment to the selected users at the last stage is less than $\alpha B$, $\alpha \in (0, 1/2]$. There might be two reasons leading to that users from $Z_2$ are not selected in $Z_2'$. The first case is when the marginal densities of some users from $Z_2$ are less than $\rho^*$, and thus we do not select them. Even if these users are all in $Z_2$, their expected total payment is at most $B/2$. Because of submodularity, the expected total loss due to these missed users is at most

$$\rho^* \cdot \frac{B}{2} = \frac{\rho_1' B}{2\delta} = \frac{V(Z_1')}{\delta}.$$

The other case is when there is not enough budget to pay for some users whose marginal densities are not less than $\rho^*$. It means that the payment for such a user (for example, user $i$) is larger than $(1/2 - \alpha)B$, i.e., $V_i(\mathcal{S})/\rho^* > (1/2 - \alpha)B$; otherwise adding this user to $Z_2'$ will not lead to that the total payment for $Z_2'$ exceeds the stage-budget $B/2$. Because $\mathbb{E}[\rho_1'] \geq \rho$, we have that

$$\mathbb{E}[V_i(\mathcal{S})] > \mathbb{E}[\rho^*] \cdot (\frac{1}{2} - \alpha)B = \frac{(1-2\alpha)\mathbb{E}[\rho_1']B}{2\delta} \geq \frac{(1-2\alpha)\rho B}{2\delta}.$$

Because the expected total payment to all users in $Z_2$ is at most $B/2$, there cannot be more than $(\frac{\delta}{1-2\alpha} - 1)$ such users in $Z_2$. Since the value of each user is at most $V(Z)/\omega$, the expected total loss due to these missed users is at most $(\frac{\delta}{1-2\alpha} - 1)V(Z)/\omega$. Therefore, we have that

$$\mathbb{E}[V(Z_2')] \geq \mathbb{E}[V(Z_2)] - (\frac{\delta}{1-2\alpha} - 1)\frac{V(Z)}{\omega} - \frac{\mathbb{E}[V(Z_1')]}{\delta}$$
$$\geq \frac{V(Z)}{2} - (\frac{\delta}{1-2\alpha} - 1)\frac{V(Z)}{\omega} - \frac{\mathbb{E}[V(Z_1')]}{\delta}$$
$$\geq [\frac{1}{2} - (\frac{\delta}{1-2\alpha} - 1)\frac{1}{\omega} - \frac{1}{\delta}]\mathbb{E}[V(Z_1')].$$

Considering both of case (a) and (b), the ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ will be at least $2\alpha/\delta$, if it satisfies that

$$\frac{1}{2} - (\frac{\delta}{1-2\alpha} - 1)\frac{1}{\omega} - \frac{1}{\delta} = \frac{2\alpha}{\delta}. \quad (1)$$

Therefore, for a specific parameter $\omega$, we can obtain the optimal ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ by solving the following optimization problem:

**Maximize** $\dfrac{2\alpha}{\delta}$ **subject to Eq. (1) and** $\alpha \in (0, 1/2]$.

When $\omega$ is sufficiently large (at least 12), we can obtain a constant ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$. Fig. 8 illustrates the optimal ratios that can be obtained by fixing proper $\delta$ when different values of $\omega$ are set. As $\omega$ becomes larger, a higher ratio can be obtained. More importantly, both the optimal ratio of $\mathbb{E}[V(Z_2')]$ to $\mathbb{E}[V(Z_1')]$ and the optimal value of $\delta$ converges fast as $\omega$ increases. Specially, the optimal ratio approaches $1/4$ as $\omega \to \infty$ and $\delta \to 4$.

### C. Proof of Lemma 9

Assume that the set of selected users computed with the budget $B'/2$ is $\mathcal{S}_l = \{1, 2, \ldots, l\}$, and the set of selected users computed with the budget $B'$ is $\mathcal{S}_k = \{1, 2, \ldots, k\}$. Then, users can be sorted according to their increasing marginal densities as follows:

$$\frac{V_1(\mathcal{S}_0)}{b_1} \geq \frac{V_2(\mathcal{S}_1)}{b_2} \geq \cdots \geq \frac{V_l(\mathcal{S}_{l-1})}{b_l} \geq \frac{2V(\mathcal{S}_l)}{B'} \geq \frac{V_{l+1}(\mathcal{S}_l)}{b_{l+1}} \geq \cdots$$
$$\geq \frac{V_k(\mathcal{S}_{k-1})}{b_k} \geq \frac{V(\mathcal{S}_k)}{B'} \geq \frac{V_{k+1}(\mathcal{S}_k)}{b_{k+1}} \geq \cdots \geq \frac{V_{|\mathcal{S}'|}(\mathcal{S}_{|\mathcal{S}'|-1})}{b_{|\mathcal{S}'|}}.$$

Thus, it can be easily derived that: $V(\mathcal{S}_l) \geq V(\mathcal{S}_k)/2$.

### D. Proof of Lemma 10

We consider two cases according to the total payment to the selected users at the last stage as follows.

**Case (a):** The total payment to the selected users at the last stage is at least $\alpha B$, $\alpha \in (0, 1/2]$. In this case, since each selected user has marginal density at least $\rho^*$, so we have that

$$V(Z_2') \geq \rho^* \alpha B = \frac{\alpha \rho_1' B}{\delta} = \frac{2\alpha V(Z_1')}{\delta}.$$

**Case (b):** The total payment to the selected users at the last stage is less than $\alpha B$, $\alpha \in (0, 1/2]$. There might be two reasons leading to that users from $Z_2$ are not selected in $Z_2'$. The first case is when the marginal densities of some users from $Z_2$ are less than $\rho^*$, and thus we do not select them. Even if these users are all in $Z_2$, their total payment is at most $B$. Because of submodularity, the total loss due to these missed users is at most

$$\rho^* \cdot B = \frac{\rho_1' B}{\delta} = \frac{2V(Z_1')}{\delta}.$$

The other case is when there is not enough budget to pay for some users whose marginal densities are not less than $\rho^*$. It means that the payment for such a user (for example, user $i$) is larger than $(1/2 - \alpha)B$, i.e., $V_i(\mathcal{S})/\rho^* > (1/2 - \alpha)B$; otherwise adding this user to $Z_2'$ will not lead to that the total

payment for $Z_2'$ exceeds the stage-budget $B/2$. Because $\rho_1' = 2V(Z_1')/B \geq V(Z)/(4B) = \rho/4$, we have that

$$V_i(\mathcal{S}) > \rho^* \cdot (\frac{1}{2} - \alpha)B = \frac{(1-2\alpha)\rho_1' B}{2\delta} \geq \frac{(1-2\alpha)\rho B}{8\delta}.$$

Because the total payment to all users in $Z_2$ is at most $B$, there cannot be more than $(\frac{8\delta}{1-2\alpha} - 1)$ such users in $Z_2$. Since the value of each user is at most $V(Z)/\omega$, the total loss due to these missed users is at most $(\frac{8\delta}{1-2\alpha} - 1)V(Z)/\omega$. Therefore, we have that

$$V(Z_2') \geq V(Z_2) - (\frac{8\delta}{1-2\alpha} - 1)\frac{V(Z)}{\omega} - \frac{2V(Z_1')}{\delta}$$
$$\geq \frac{V(Z)}{4} - (\frac{8\delta}{1-2\alpha} - 1)\frac{V(Z)}{\omega} - \frac{2V(Z_1')}{\delta}$$
$$\geq [\frac{1}{4} - (\frac{8\delta}{1-2\alpha} - 1)\frac{1}{\omega} - \frac{2}{\delta}]V(Z_1').$$

Considering both of case (a) and (b), the ratio of $V(Z_2')$ to $V(Z_1')$ will be at least $2\alpha/\delta$, if it satisfies that

$$\frac{1}{4} - (\frac{8\delta}{1-2\alpha} - 1)\frac{1}{\omega} - \frac{2}{\delta} = \frac{2\alpha}{\delta}. \quad (2)$$

Therefore, for a specific parameter $\omega$, we can obtain the optimal ratio of $V(Z_2')$ to $V(Z_1')$ by solving the following optimization problem:

**Maximize** $\dfrac{2\alpha}{\delta}$ **subject to Eq. (2) and** $\alpha \in (0, 1/2]$.

When $\omega$ is sufficiently large, we can obtain a constant ratio of $V(Z_2')$ to $V(Z_1')$. Specially, the optimal ratio approaches $1/12$ as $\omega \to \infty$ and $\delta \to 12$.

### E. Proof of Lemma 13

Consider a user $i$ with true type $\theta_i = (a_i, d_i, \Gamma_i, c_i)$, and reported strategy type $\hat{\theta}_i = (\hat{a}_i, \hat{d}_i, \Gamma_i, b_i)$. According to the *OMG* mechanism, at each time step $t \in [\hat{a}_i, \hat{d}_i]$, there may be a new decision on whether to accept user $i$, and at what price. For convenience, let $T_t'$, $B_t'$, $\rho_t^*$, and $\mathcal{S}_t$ denote the end time of the current stage, the residual budget, the current density threshold, and the set of selected users respectively at time step $t$ and before making decision on user $i$. Let $\hat{\theta}_{-i}$ denote the strategy types of all users excluding $\hat{\theta}_i$. We first prove the following two propositions.

*Proposition (a): at some time step $t \in [\hat{a}_i, \hat{d}_i]$, fix $\rho_t^*$ and $B_t'$, reporting the true cost is a dominant strategy for user $i$.* It can be easily proved since the decision at time step $t$ is bid-independent.

*Proposition (b): fix $b_i$ and $\hat{\theta}_{-i}$, reporting the true arrival/departure time is a dominant strategy for user $i$.* It's because that user $i$ is always paid for a price equal to the maximum price attained during his/her reported arrival-departure interval. Assume that user $i$ can obtain the maximum payment at time step $t \in [\hat{a}_i, \hat{d}_i]$. Then reporting an earlier arrival time or a later departure time than $t$ does not affect the payment of user $i$. However, if user $i$ reports a later arrival time or an earlier departure time than $t$, then he/she will obtain a lower payment.

Based on the proposition (b), it is sufficient to prove this lemma by adding a third proposition:

*Proposition (c): fix $[a_i, d_i]$ and $\hat{\theta}_{-i}$, reporting the true cost is a dominant strategy for user $i$.* According to the proposition (a), reporting a false cost at time step $t$ cannot improve user $i$'s payment at the current time. Thus, it only needs to prove that *reporting a false cost at time step $t \in [a_i, d_i)$ still cannot improve user $i$'s payment at time step $t'(t < t' \leq d_i)$.*

First, we consider the case when user $i$ is selected as a winner by reporting his/her true type at time step $t = a_i$. In this case he/she satisfies $b_i \leq V_i(\mathcal{S}_t)/\rho_t^* \leq B_t'$, and he/she can obtain the payment $V_i(\mathcal{S}_t)/\rho_t^*$. At time $t'(t < t' < T_t')$, due to the submodularity of $V(\mathcal{S})$, we have $V_i(\mathcal{S}_{t'}) \geq V_i(\mathcal{S}_t)$. Then user $i$ will obtain the payment $V_i(\mathcal{S}_{t'})/\rho_t^*$ if $b_i \leq V_i(\mathcal{S}_{t'})/\rho_t^* \leq B_{t'}'$, otherwise he/she will obtain the payment 0. Thus, user $i$ cannot obtain higher payment at time step $t'$ than that at $t$. It means that a user cannot improve his/her payment by reporting a false cost if his/her arrival-departure interval does not span more than one stage.

Next we consider user $i$'s payment at time step $t'(T_t' \leq t' \leq d_i)$ if his/her arrival-departure interval spans multiple stages. According to the proposition (a), user $i$'s payment at time step $t'$ depends on $\rho_{t'}^*$ and $B_{t'}'$. Because $\rho_{t'}^*$ is independent with $b_i$, it only needs to consider the effect of $b_i$ on $B_{t'}'$. If user $i$ reports a false cost $b_i$ which still satisfies $b_i \leq V_i(\mathcal{S}_t)/\rho_t^* \leq B_t'$, then he/she is still accepted at price $V_i(\mathcal{S}_t)/\rho_t^*$ at time step $t$, and thus $B_{t'}'$ remains unchanged. If user $i$ reports a larger bid $b_i > c_i$ and $b_i > V_i(\mathcal{S}_t)/\rho_t^*$, then he/she will not selected at time step $t$. In this case, more budget will be allocated for other users, and $B_{t'}'$ will be diminished. Therefore, user $i$ cannot obtain higher payment at time step $t'$.

Second, we consider the case when user $i$ is not selected as a winner by reporting his/her true type at time step $t = a_i$. In this case it satisfies $c_i > V_i(\mathcal{S}_t)/\rho_t^*$, or $V_i(\mathcal{S}_t)/\rho_t^* > B_t'$. In case $c_i > V_i(\mathcal{S}_t)/\rho_t^*$, if user $i$ reports a false cost $b_i$ which still satisfies $b_i > V_i(\mathcal{S}_t)/\rho_t^*$, then the outcome remains unchanged. If user $i$ reports a lower bid $b_i < c_i$ and $b_i \leq V_i(\mathcal{S}_t)/\rho_t^*$, then he/she will be accepted at price $V_i(\mathcal{S}_t)/\rho_t^*$ at time step $t$. In such case, however, his/her utility will be negative. In addition, $B_{t'}'$ remains unchanged, and thus user $i$'s payment at time step $t' > t$ is not affected. In case $V_i(\mathcal{S}_t)/\rho_t^* > B_t'$, reporting a false cost does not affect the outcome at time step $t$ or the residual budget $B_{t'}'$ at time step $t' > t$. To sum up, reporting a false cost cannot improve user $i$'s payment at time step $t' > t$.

## REFERENCES

[1] G. Chatzimilioudis, A. Konstantinidis, C. Laoudias, and D. Zeinalipour-Yazti, "Crowdsourcing with smartphones," *IEEE Internet Comp.*, pp. 36–44, 2012.

[2] "Sensorly," http://www.sensorly.com.

[3] E. Koukoumidis, L.-S. Peh, and M. R. Martonosi, "Signalguru: leveraging mobile phones for collaborative traffic signal schedule advisory," in *Proc. ACM MobiSys*, 2011, pp. 127–140.

[4] P. Mohan, V. N. Padmanabhan, and R. Ramjee, "Nericell: rich monitoring of road and traffic conditions using mobile smartphones," in *Proc. ACM SenSys*, 2008, pp. 323–336.

[5] A. Thiagarajan, L. Ravindranath, K. LaCurts, S. Madden, H. Balakrishnan, S. Toledo, and J. Eriksson, "Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones," in *Proc. ACM SenSys*, 2009, pp. 85–98.

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

[7] M. Stevens and E. DHondt, "Crowdsourcing of pollution data using smartphones," in *Proc. ACM UbiComp*, 2010, pp. 1–4.

[8] N. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, and A. Campbell, "A survey of mobile phone sensing," *IEEE Commun. Mag.*, vol. 48, no. 9, pp. 140–150, 2010.

[9] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges," *IEEE Commun. Mag.*, vol. 49, no. 11, pp. 32–39, 2011.

[10] G. Danezis, S. Lewis, and R. Anderson, "How much is location privacy worth," in *Proc. WEIS*, 2005.

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

[12] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa, and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing," in *Proc. IEEE INFOCOM*, 2012, pp. 1701–1709.

[13] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: incentive mechanism design for mobile phone sensing," in *Proc. ACM MobiCom*, 2012, pp. 173–184.

[14] L. Jaimes, I. Vergara-Laurens, and M. Labrador, "A location-based incentive mechanism for participatory sensing systems with budget constraints," in *Proc. IEEE PerCom*, 2012, pp. 103–108.

[15] Z. Feng, Y. Zhu, Q. Zhang, L. M. Ni, and A. V. Vasilakos, "Trac: Truthful auction for location-aware collaborative sensing in mobile crowdsourcing," in *Proc. IEEE INFOCOM*, 2014, pp. 1231–1239.

[16] T. Luo, H.-P. Tan, and L. Xia, "Profit-maximizing incentive for participatory sensing," in *Proc. IEEE INFOCOM*, 2014, pp. 127–135.

[17] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proc. WWW*, 2013, pp. 1157–1166.

[18] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proc. WWW*, 2013, pp. 1167–1178.

[19] A. Badanidiyuru, R. Kleinberg, and Y. Singer, "Learning on a budget: posted price mechanisms for online procurement," in *Proc. ACM EC*, 2012, pp. 128–145.

[20] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems," in *Proc. IEEE INFOCOM*, 2013, pp. 1402–1410.

[21] M. T. Hajiaghayi, R. Kleinberg, and D. C. Parkes, "Adaptive limited-supply online auctions," in *Proc. ACM EC*, 2004, pp. 71–80.

[22] M. Babaioff, N. Immorlica, D. Kempe, and R. Kleinberg, "Online auctions and generalized secretary problems," *ACM SIGecom Exchanges*, vol. 7, no. 2, pp. 1–11, 2008.

[23] Z. Bar-Yossef, K. Hildrum, and F. Wu, "Incentive-compatible online auctions for digital goods," in *Proc. ACM-SIAM SODA*, 2002, pp. 964–970.

[24] M. Bateni, M. Hajiaghayi, and M. Zadimoghaddam, "Submodular secretary problem and extensions," in *APPROX-RANDOM*, 2010, pp. 39–52.

[25] A. V. Goldberg, J. D. Hartline, A. R. Karlin, M. Saks, and A. Wright, "Competitive auctions," *Games and Economic Behavior*, vol. 55, no. 2, pp. 242–269, 2006.

[26] Y. Singer, "Budget feasible mechanisms," in *Proc. IEEE FOCS*, 2010, pp. 765–774.

[27] S. Khullera, A. Mossb, and J. Naor, "The budgeted maximum coverage problem," *Information Processing Letters*, vol. 70, pp. 39–45, 1999.

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

[29] K. Lee, S. Hong, S. Kim, I. Rhee, and S. Chong, "Slaw: A mobility model for human walks," in *Proc. IEEE INFOCOM*, 2009, pp. 855–863.

[30] D. Zhao, H. Ma, and L. Liu, "Frugal online incentive mechanisms for crowdsourcing tasks truthfully," Technical Report, 2014. [Online]. Available: http://arxiv.org/abs/1404.2399

[31] X. Zhang, Z. Yang, Z. Zhou, H. Cai, L. Chen, and X. Li, "Free market of crowdsourcing: Incentive mechanism design for mobile sensing," *IEEE Trans. on Parallel and Distributed Systems*, 2014.

[32] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Proc. Pervasive*, 2010, pp. 138–155.

[33] M. Babaioff, N. Immorlica, D. Kempe, and R. Kleinberg, "A knapsack secretary problem with applications," in *APPROX-RANDOM*, 2007, pp. 16–28.

[34] R. Kleinberg, "A multiple-choice secretary algorithm with applications to online auctions," in *Proc. ACM-SIAM SODA*, 2005, pp. 630–631.