SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

December 2021, Vol. 64 222106:1–222106:15
https://doi.org/10.1007/s11432-019-2730-8

Truthfully coordinating participation routes in informative participatory sensing

Shaofei CHEN[1*], Dengji ZHAO[2], Alexandros ZENONOS[3],
Jing CHEN[1] & Lincheng SHEN[1]

[1]*College of Intelligence Science and Technology, National University of Defense Technology, Changsha* 410073*, China;*
[2]*School of Information Science and Technology, ShanghaiTech University, Shanghai* 201210*, China;*
[3]*School of Electronics and Computer Science, University of Southampton, Southampton* SO17 1BJ*, UK*

Received 29 May 2019/Revised 10 September 2019/Accepted 8 November 2019/Published online 25 November 2021

Abstract Participatory sensing is a promising approach with which people contribute sensory information to form a body of knowledge. In practice, people may have different ways to engage in a participatory sensing campaign. For example, there are several possible routes from a participant's home to her office, where a route can be seen as a set of space-temporal coordinates, and measurements can be taken at these coordinates. To coordinate participation routes to collect more valuable information with a limited number of participants, a further concept, informative participatory sensing (IPS) has been developed recently. However, existing IPS systems lack incentive mechanisms to fight against the strategic behaviours of self-interested users. Hence, we propose a formal model of IPS where a service provider can coordinate individual schedules of self-interested participants. As the problem of informative path coordination is NP-hard, the well-known mechanism, Vickrey-Clarke-Groves (VCG) will be computationally inefficient to solve our problem. Given this, we design a sequentially sorting mechanism (SSM) for the model to allocate the schedules and determine the bonuses for these participants, and we then theoretically prove that SSM is computationally efficient, individually rational, profitable and truthful. Furthermore, we empirically evaluate our route allocation method in simulations and show that it significantly outperforms several benchmark approaches.

Keywords coordinating participation routes, informative participatory sensing, incentive mechanism, computationally efficient

Citation Chen S F, Zhao D J, Zenonos A, et al. Truthfully coordinating participation routes in informative participatory sensing. Sci China Inf Sci, 2021, 64(12): 222106, https://doi.org/10.1007/s11432-019-2730-8

1 Introduction

Participatory sensing [1] provides a new architecture of human carrying portable or wearable devices to gather their local knowledge and send it to service providers. These devices, such as smartphones, are programmable and equipped with a set of cheap but powerful embedded sensors, such as cameras, GPS and digital compass. Examples of applications include Waze [2] (taking live traffic information from users on the road for traffic estimation), Weathermob [3] (collecting real-time weather reports to improve the accuracy of weather forecasting), NoiseSPY [4] (collecting noise data for urban noise monitoring and mapping) and Ushahidi [5] (aggregating people's reports such as text messages and photographs across a disaster zone). Compared to the traditional approach of using a large number of fixed sensors, participatory sensing reduces the fundamental cost of building fixed sensing systems and has achieved impressive results in applications [6, 7]. Moreover, some incentive mechanisms for participatory sensing systems are designed for stimulating user participation [8–12].

However, most existing participatory campaigns lack a coordination element to guide participants' possible ways of engaging in. This is important because the data obtained from the campaign are usually limited, and the system with a coordination approach may acquire better situational awareness [13]. For example, each participant may have a set of different routes that access her destination and participation

* Corresponding author (email: chenshaofei01@nudt.edu.cn)

with coordinated routes can improve the quality of the data to be collected. Recent studies [14–16] first studied the problems of coordinating participants' sensing behaviours in participatory sensing. We call this type of problems informative participatory sensing (IPS), in which the platform guides or suggests participants where and when to take measurements to fill coverage gaps throughout time in the area of interest. Moreover, to the best of our knowledge, no incentive mechanism for the IPS problems has been designed.

In this paper, we focus on a class of IPS problems where the service provider can coordinate participants' schedules to maximize the information collected while mitigating the costs incurred. In particular, we expect to design an incentive mechanism to motivate and coordinate self-interested users to engage in. Specifically, a participant's routine route can be seen as a set of space-temporal coordinates and measurements can be taken at these coordinates with an amount of cost. Given this, each participant reports her profile that includes a set of candidate routes, and the mechanism then computes an allocation specifying who is going to take measurements along which of her candidate routes and determines to pay how much bonus to each participant. Generally, a participant may like to partially conceal their candidate routes if this may influence the routes allocation or bonus determination to get more utility. The goal of an IPS system is then to maximize social welfare while stimulating participation and truthful reporting.

Against this background, we propose a new model for IPS and design a mechanism to compute the allocations and bonuses for participants. In more detail, we first extend the IPS model in [14] to feature the situations where a service provider can coordinate individual schedules of self-interested participants. Then, we propose a Vickrey-Clarke-Groves (VCG) [17] based on the mechanism for IPS. As the problem of informative path coordination is known to be non-deterministic polynomial-time-hard (NP-hard) [18], the VCG mechanism will be computationally inefficient to solve IPS. Given this, we design a sequentially sorting mechanism (SSM) and prove that it is individually rational, truthful, profitable, and computationally efficient. This paper advances the state-of-the-art in the following ways.

- We propose a new formal model of IPS where a service provider can coordinate individual schedules of self-interested participants. The model not only features spatio-temporal dynamics in large environments and the coordination to guide participants' possible ways of engaging in, but also takes into account that participants are self-interested and may partially conceal their candidate routes.
- We design a mechanism, SSM, for the IPS model, which is proven to be computationally efficient, individually-rational, profitable and truthful.
- Furthermore, we empirically evaluate our allocation method in simulations and show that it significantly outperforms several benchmark approaches.

The remainder of the paper is organised as follows. First, we present the related studies in Section 2 and introduce the IPS framework in Section 3. We then define the model of IPS in Section 4 and design the mechanisms for the model in Section 5. Next, we present the performance evaluations in Section 6 and finally conclude this paper in Section 7.

2 Related work

In this section, we review related work on incentive mechanisms for participatory sensing and multi-agent information gathering.

2.1 Incentive mechanisms for participatory sensing

Mobile crowdsensing is an emerging and powerful paradigm that takes advantage of portable devices to collect data beyond the scale of what was previously possible [19, 20]. As these devices are equipped with various sensors that are ubiquitous in people everyday lives, mobile crowdsensing is enabling a wide variety of applications like transportation monitoring [2], weather forecast [3], environmental monitoring [4], health monitoring [21] and safety [5]. In accordance with the awareness and involvement of people, mobile crowdsensing can be categorised into two major classes [19]: participatory sensing and opportunistic sensing. Participatory sensing requires the active involvement of participants to contribute sensor data (e.g., taking a picture, reporting a victim's location). Opportunistic sensing is more autonomous, and user involvement is minimal (e.g., continuous location sampling without the explicit action of the user). We refer the reader to [19, 20] for more details on mobile crowdsensing.

Although participatory sensing tools have proven technically feasible, the problem of motivating potential observers to collect data is still challenging. Hence, designing incentive mechanisms for stimulating participatory sensing applications is a hot research issue [22]. Most related to our IPS, Refs. [9, 12, 23] develop mechanisms for fusing environmental information using participatory sensing. The information value of observations taken by participants is cast as a submodular function in these studies and they address the questions of how to optimally choose the participants from a large population of strategic users and compensate them for their participation. These studies assume that participants strategize about the costs of taking observations [9, 12] or the number of observations they can take [23]. However, none of these studies consider that each participant may have several possible ways to engage in and participants may have different candidate ways.

2.2 Multi-agent information gathering

In general, methods to gather situational awareness are typically categorised as a class of information gathering problems [24–26], in which agents aim to collect and provide up-to-date situational awareness of the environmental phenomena. The information gathering tasks of agents can be formulated as identifying a set of observations that maximize an observation value function quantifying their informativeness. Gaussian processes (GPs) [27] are often used to model the underlying phenomenon, and many metrics used to quantify the amount of information collected satisfy submodularity [28], an intuitive diminishing returns property. Ref. [24] proved that planning such agents' informative paths is an NP-hard optimisation problem and proposed a sequential allocation method to near-optimally solve this problem. The algorithm designed in [25] enables agents to coordinate their movements with their direct neighbours to maximize the collective information gain. For continuously gathering information in such environments, a near-optimal multi-agent algorithm was developed in [26].

A recent study [14] first designed a method for coordinating the measurements by using a large group of people to gather information in participatory sensing. The authors then extended the model to feature the uncertainty that people may or may not actually take a suggested measurement when requested to do so by the system and proposed a stochastic coordination algorithm [16] to solve this model. Moreover, their study in [29] designed a trust-based algorithm to effectively coordinate measurements in the presence of malicious user behaviour. However, these methods for multi-agent information gathering problems are not sufficient to fight against strategic behaviours of self-interested users. Hence, we next study incentive mechanisms for the IPS problems in this paper.

3 The IPS framework

This section introduces the framework of IPS for environmental monitoring. In particular, we explain the components of an IPS system and present the process of running an IPS campaign.

In the IPS system, as shown in Figure 1, there is an IPS platform and a large group of participants with sensing devices. The platform and the participants are connected via cloud. The environmental phenomenon is modelled in the environmental representation component and trained by historic sensed data. The participants report their candidate routes to the platform, while the costs of these routes are calculated by the cost estimation component and known to both the platform and the participants. The key component is the IPS mechanism, which coordinates the participants' routes by computing an allocation specifying who is going to take measurements along which of her candidate routes and determines to pay how much bonus to each participant. The participants then collect the environmental data along their allocated routes and send these data to the platform.

Given this framework, Sections 4 and 5 formulate and solve the coordination problem of IPS.

4 Coordination in IPS

In this section, we formally model the coordination problem in IPS, where the information gathering campaign is seen as participants collecting a set of observations in city environments. Specifically, we first formulate the participant and the environment, and then go on to model the mechanism by which the participants are engaging in IPS campaigns.

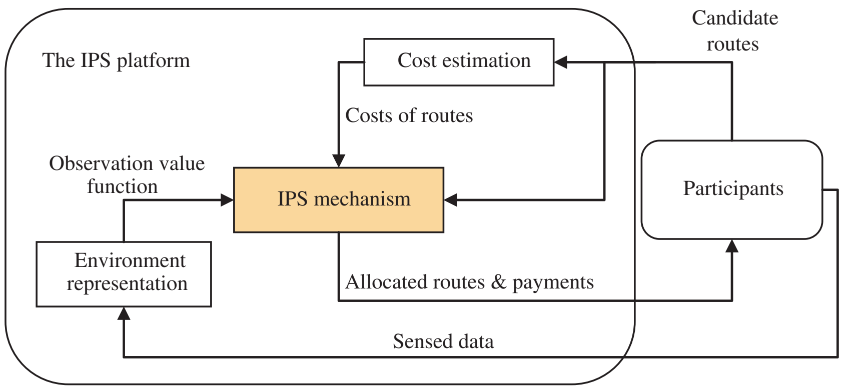

Figure 1 (Color online) A conceptual architecture of IPS.

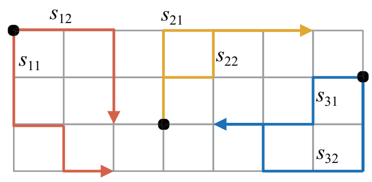

Figure 2 (Color online) An example of three agents' candidate routes.

4.1 Participant

The participant is modelled by her way of taking observations in the environment. In particular, a participant can take observations along her routine route with a portable sensing device. These observations will lead to better situational awareness with a cost.

Definition 1. An agent (i.e., a participant), $i \in N$ is a user who carries a portable device and collects information with their sensing measurements.

Definition 2. O is a set of spatio-temporal observation coordinates and an observation $o \in O$ denotes when and where measurements can be taken by some agents. In more detail, we have $O = (L \cup \{\perp\}) \times T$ where L is a set of spatial coordinates where measurements can be taken and T is a discrete set of temporal coordinates that specify when measurements can be taken.

Definition 3. Each agent i has a privately observed type, $\theta_i = \{s_{ir}\}_{r \in R}$, where $s_{ir} \subseteq O$, a set of observation coordinates that i may come across, represents a routine route of i may choose and R is the set of sequence numbers of her candidate routes.

Figure 2 shows an example of three agents' candidate routes, where each agent has reported two different routes that access to her destination. In practice, an agent can use virtual travel assistants to select and report her candidate routes.

Next, we continue to define the cost and the value of taking observations.

Definition 4. A cost function $c : 2^O \rightarrow \mathbb{R}^+$, indicates the cost for an agent to take a set of observations. $c(s_{ir})$ is the cost of agent i taking observations s_{ir} and $c(\emptyset) = 0$.

We assume that the costs of taking observations along different routes are independent of each other. In other words, an observation cost only directly depends on the features of its routes instead of other routes. Moreover, we assume that the costs of all routes can be calculated (or estimated) by the system and publicly known. For example, costs of routes normally can be calculated from their lengths and the traffic congestion degrees. Our setting could be more general by allowing each agent to have their own cost function, but the cost functions still need to be known to the mechanism. In addition, costs and available routes could also both be private, but agents only strategize about which routes they report.

Definition 5. An observation value function $f : 2^O \rightarrow \mathbb{R}^+$, assigns an observation value to a set of observations, where the observation value quntifies the increase in situational awareness that these observations bring about.

Here we present the properties of the observation value by summarising different metrics of the value of collected information. These metrics include mutual information [30], entropy [31] and informative path [18], and the properties which they have in common are described as follows.

First, observation value functions are non-decreasing: $\forall \Phi \subseteq \Phi' \subseteq O, f(\Phi) \leqslant f(\Phi')$. That is, taking more observations never "lose" situational awareness.

Second, observation value functions are submodular: $\forall \Phi \subseteq \Phi' \subseteq O$ and $\forall o \in O$:

$$f\left(\Phi \cup \{o\}\right) - f\left(\Phi\right) \geqq f\left(\Phi' \cup \{o\}\right) - f\left(\Phi'\right).$$

This is an intuitive property of diminishing returns—making an additional observation helps more if we have made only a few observations so far, and less if we have made many observations. Another equivalent condition is $\forall \Phi, \Phi' \subseteq O$:

$$f\left(\Phi\right) + f\left(\Phi'\right) \geqq f\left(\Phi \cup \Phi'\right) + f\left(\Phi \cap \Phi'\right). \tag{1}$$

Given this, we can derive that the observation value of a route also satisfies submodularity as follows: $\forall \Phi \subseteq \Phi' \subseteq O$, $s \subseteq O$ and $i \in N$: $f(\Phi \cup s) - f(\Phi) \geqq f(\Phi' \cup s) - f(\Phi')$.

We next specifically propose how we model the environment and use it to compute the observation value.

4.2 Environment

For IPS, the observation data obtained by all participants are used to depict the situation of the environment. One of the key challenges is to use these limited data to predict its state at unobserved coordinates. A recent study has addressed this challenge by modelling the environmental phenomenon as GPs [27], which is an effective model that captures the relationship of different spatio-temporal coordinates with respect to the environmental phenomenon. Hence, we use GP to model the environmental phenomenon.

An important property of the GP is that for every finite subset A of the observations O, the joint distribution over the corresponding random variables $\boldsymbol{x}_A$ is Gaussian. In more detail, the GP is specified by a mean $m(\boldsymbol{v})$ and a covariance function (also known as kernel) $k(\boldsymbol{v}, \boldsymbol{v}')$. Formally, $f(\boldsymbol{v}) = \mathcal{GP}(m(\boldsymbol{v}), k(\boldsymbol{v}, \boldsymbol{v}'))$. Let A denote the matrix with the coordinates at which observations were made, and let vector v_A denote the observed value at those coordinates. Then, the predictive distribution of the observation at coordinates y is Gaussian with mean $\mu_{y|A}$ and variance $\Sigma_{y|A}$, which is given by

$$\begin{aligned} \mu_{y|A} &= \mu_y + \Sigma_{yA}\Sigma_{AA}^{-1}(v_A - \mu_A), \\ \Sigma_{y|A} &= \Sigma_{yy} - \Sigma_{yA}\Sigma_{AA}^{-1}\Sigma_{Ay}. \end{aligned} \tag{2}$$

Following the selection in [14, 16], we adopt Matérn as the covariance function:

$$k(\boldsymbol{v}, \boldsymbol{v}') = \sigma_f^2(1 + \sqrt{3}r)\exp(-\sqrt{3}r) + \sigma_n^2\delta_{\boldsymbol{v},\boldsymbol{v}'}, \tag{3}$$

where

$$r = \sqrt{(\boldsymbol{v} - \boldsymbol{v}')^{\mathrm{T}}\boldsymbol{P}^{-1}(\boldsymbol{v} - \boldsymbol{v}')}, \quad \boldsymbol{P} = \begin{bmatrix} l_1 & 0 & 0 \\ 0 & l_2 & 0 \\ 0 & 0 & l_3 \end{bmatrix},$$

and $\theta = \{l_1, l_2, l_3, \sigma_f^2, \sigma_n^2\}$ are the hyperparameters that need to be learned. l_1, l_2, l_3 capture the dynamism of the phenomenon in both spatial and temporal dimension, σ_f^2 and σ_n^2 are parameters that control the sensitivity of the kernel to both measurements and noise while $\delta_{\boldsymbol{v},\boldsymbol{v}'}$ is the Kronecker delta, which is 1 if $\boldsymbol{v} = \boldsymbol{v}'$, and 0 if $\boldsymbol{v} \neq \boldsymbol{v}'$.

Given this, the GP provides the mathematics of computing the observation value. For a set of observations, their observation value can be seen as proportional to the entropy without making any observations minus the uncertainty when these observations are made.

Having modelled the participants' behaviours and the environment, we next model the mechanism.

4.3 Mechanism

In a participatory sensing campaign, each agent i is required to report her type to the system. We denote the report by $\hat{\theta}_i$ and note that agents could misreport their type if it is their best interest to do so. Let $\theta = (\theta_i)_{i \in N}$ be the type profile of all agents, $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$ be the type profile of all agents except i. Given agent i of type θ_i, we refer to Θ_i as the set of all possible type reports of i, and let Θ be the set of all possible type profile reports of all agents with type profile θ.

Assumption 1. We assume that any agent i can only report a profile $\hat{\theta}_i$ that satisfies $\hat{\theta}_i \subseteq \theta_i$, which means that an agent can only report some of her feasible candidate routes. Otherwise, some agent may get an infeasible allocation from the system.

Given a report $\hat{\theta}$, an IPS mechanism will compute an allocation (i.e., schedules) $\pi = \{\pi_i(\hat{\theta})\}_{i\in N}$ and bonuses $x = \{x_i(\hat{\theta})\}_{i\in N}$, where an allocated route $\pi_i(\hat{\theta}) \in \hat{\theta}_i \cup \emptyset$ is the route for i to engage in the campaign, a bonus $x_i(\hat{\theta})$ is used to reward the participation of i. $\pi_i(\hat{\theta}) = \emptyset$ means no route is allocated. To note, compared with the model (proposed in [14]) in which an agent selects several points from her route to take observations, in our setting an agent takes observations at all positions along one of her candidate routes.

We next define the social welfare, the profit of the platform and the utility of an agent, respectively.

Definition 6. The social welfare of an allocation $\pi(\hat{\theta}) = \{\pi_i(\hat{\theta})\}_{i\in N}$ is the observation value obtained by using this allocation minus the costs incurred:

$$\mathrm{SW}(\pi(\hat{\theta})) = f(\cup_{i\in N}\pi_i(\hat{\theta})) - \sum_{i\in N} c(\pi_i(\hat{\theta})).$$

The observation value in this social welfare function can be computed with GPs model as presented in Subsection 4.2 and we will discuss the computation efficiency of finding optimal allocation later in Lemma 1 and Theorem 4.

Definition 7. The profit of the IPS platform is the observation value obtained minus the sum of participants' bonuses:

$$u(\hat{\theta}, (\pi, x)) = f(\cup_{i\in N}\pi_i(\hat{\theta})) - \sum_{i\in N} x_i(\hat{\theta}).$$

Definition 8. Given agents' type profile θ, their reports $\hat{\theta}$ and mechanism $\mathcal{M} = (\pi, x)$, we define the utility of i as

$$u(\theta_i, \hat{\theta}, (\pi, x)) = x_i(\hat{\theta}) - c(\pi_i(\hat{\theta})), \tag{4}$$

where $c(\pi_i(\hat{\theta}))$ is the cost of taking observations along the route $\pi_i(\hat{\theta})$.

Given this, our goal is to find an incentive mechanism satisfying the following four desirable properties.

- Individually rational. An agent never pays more than what she gets.
- Truthful. Truthfulness ensures that agents cannot get more utility by reporting a false profile (i.e, concealing some candidate routes). That is, $u(\theta_i, (\theta_i, \hat{\theta}_{-i}), (\pi, x)) \geqslant u(\theta_i, \hat{\theta}, (\pi, x))$ for all $i \in N$, all type profile $\theta \in \Theta$, all reported profile $\hat{\theta} \in \Theta$, i.e., rational agents will truthfully report all of their candidate routes.
- Profitable. The system should not run the mechanism with a deficit, i.e., the total bonus paid is no more than the value of gathered information.
- Computationally efficient. The allocation $\pi(\hat{\theta}) = \{\pi_i(\hat{\theta})\}_{i\in N}$ and bonuses $x(\hat{\theta}) = \{x_i(\hat{\theta})\}_{i\in N}$ can be computed in polynomial time.

Having formulated the coordination problem in IPS, we next design two mechanisms to allocate the schedules and determine the payments of the participants.

5 Mechanisms design

This section introduces the VCG mechanism and the SSM mechanism, respectively.

5.1 VCG Mechanism

In this subsection, we first propose the VCG mechanism (as shown in Algorithm 1) and then analyse its properties.

Given reported profile $\hat{\theta}$, VCG mechanism [32] efficiently allocates routes to agents (lines 3–5), where the efficient allocation is defined as follows.

Definition 9. An allocation π is efficient, if for all $\theta \in \Theta$, $\pi(\theta) = \arg\max_{\pi'(\theta)\in\Pi} \mathrm{SW}(\pi'(\theta))$, where $\Pi = \prod_{i\in N} \hat{\theta}_i \cup \emptyset$ is the set of all possible allocations.

Algorithm 1 The VCG mechanism

1: **Input**: Report of candidate routes $\hat{\theta} = (\hat{\theta})_{i \in N}$.
2: **Output**: A schedule π and bonuses x.
{Phase 1: Route allocation.}
3: $\Pi \leftarrow \prod_{i \in N} \hat{\theta}_i \cup \emptyset$;
4: $\pi(\hat{\theta}) \leftarrow \arg\max_{\pi'(\theta) \in \Pi} \mathrm{SW}(\pi'(\theta))$;
5: $\pi \leftarrow \pi(\hat{\theta})$;
{Phase 2: Bonus determination.}
6: **for** $i \in N$ **do**
7: $\quad \mathrm{SW}_{-i}(\pi(\hat{\theta})) = f(\cup_{j \in N} \pi_j(\hat{\theta})) - \sum_{j \in N \setminus \{i\}} c(\pi_j(\hat{\theta}))$;
8: $\quad \Pi \leftarrow \prod_{j \in N \setminus \{i\}} \hat{\theta}_j \cup \emptyset$;
9: $\quad \mathrm{SW}(\pi(\hat{\theta}_{-i})) \leftarrow \max_{\pi'(\theta_{-i}) \in \Pi} \mathrm{SW}(\pi'(\theta_{-i}))$;
10: $\quad x_i^{\mathrm{Clarke}}(\hat{\theta}) \leftarrow \mathrm{SW}_{-i}(\pi(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i}))$;
11: **end for**
12: $x \leftarrow \{x_i^{\mathrm{Clarke}}(\hat{\theta})\}_{i \in N}$;
13: **return** (π, x).

Note that in this paper, efficient allocation means it maximizes the social welfare, while mechanism's computationally efficient means the allocation and bonuses can be computed in polynomial time.

Given reported profile $\hat{\theta}$ and an efficient allocation π, the bonus for agent i is calculated by (line 10)

$$x_i^{\mathrm{Clarke}}(\hat{\theta}) = \mathrm{SW}_{-i}(\pi(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i})), \tag{5}$$

where $\mathrm{SW}_{-i}(\pi(\hat{\theta})) = f(\cup_{j \in N} \pi_j(\hat{\theta})) - \sum_{j \in N \setminus \{i\}} c(\pi_j(\hat{\theta}))$, i.e., the social welfare of an efficient allocation excluding the value (i.e., cost) of agent i, and $\mathrm{SW}(\pi(\hat{\theta}_{-i})) = f(\cup_{j \in N \setminus \{i\}} \pi_j(\hat{\theta}_{-i})) - \sum_{j \in N \setminus \{i\}} c(\pi_j(\hat{\theta}_{-i}))$, i.e., the social welfare of an efficient allocation when excluding agent i's report from the report of all agents.

We next analyse the properties of the VCG mechanism.

Theorem 1. VCG mechanism is individually rational.

Proof. The utility of agent i satisfies $u(\theta_i, \hat{\theta}, (\pi, x)) = x_i^{\mathrm{Clarke}}(\hat{\theta}) - c(\pi_i(\hat{\theta})) = \mathrm{SW}_{-i}(\pi(\hat{\theta})) - c(\pi_i(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i})) = \mathrm{SW}(\pi(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i})) \geqslant 0$ where the inequality follows from that $\forall i \in N, \pi(\hat{\theta}_{-i})$ can be seen as one of the allocations given $\hat{\theta}$ while $\pi(\hat{\theta})$ is an efficient allocation. Thus VCG mechanism is individually rational.

Theorem 2. VCG mechanism is truthful.

Proof. We first have that $u(\theta_i, (\theta_i, \hat{\theta}_{-i}), (\pi, x)) = x_i^{\mathrm{Clarke}}(\theta_i, \hat{\theta}_{-i}) - c(\pi_i(\theta_i, \hat{\theta}_{-i})) = \mathrm{SW}(\pi(\theta_i, \hat{\theta}_{-i})) - \mathrm{SW}(\pi(\hat{\theta}_{-i}))$ and $u(\theta_i, \hat{\theta}, (\pi, x)) = x_i^{\mathrm{Clarke}}(\hat{\theta}) - c(\pi_i(\hat{\theta})) = \mathrm{SW}(\pi(\hat{\theta}_i)) - \mathrm{SW}(\pi(\hat{\theta}_{-i}))$. Given Assumption 1, we know that $\theta_i \supseteq \hat{\theta}_i$, and then any allocation from $(\hat{\theta}_i, \hat{\theta}_{-i}) = \hat{\theta}$ must also be a possible allocation from $(\theta_i, \hat{\theta}_{-i})$. As π is efficient, we can get that $\mathrm{SW}(\pi(\theta_i, \hat{\theta}_{-i})) \geqslant \mathrm{SW}(\pi(\hat{\theta}))$ and then $u(\theta_i, (\theta_i, \hat{\theta}_{-i}), (\pi, x)) \geqslant u(\theta_i, \hat{\theta}, (\pi, x))$. Thus we have proven that this VCG mechanism is truthful.

Theorem 3. VCG mechanism is profitable.

Proof. The profit of the platform satisfies

$$\begin{aligned}
& u(\hat{\theta}, (\pi, x)) = f(\cup_{i \in N} \pi_i(\hat{\theta})) - \sum_{i \in N} x_i^{\mathrm{Clarke}}(\hat{\theta}) \\
=& \mathrm{SW}(\pi(\hat{\theta})) - \sum_{i \in N} \Big(x_i^{\mathrm{Clarke}}(\hat{\theta}) - c(\pi_i(\hat{\theta})) \Big) \\
=& \mathrm{SW}(\pi(\hat{\theta})) - \sum_{i \in N} \Big(\mathrm{SW}_{-i}(\pi(\hat{\theta})) - c(\pi_i(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i})) \Big) \\
=& \mathrm{SW}(\pi(\hat{\theta})) - \sum_{i \in N} \Big(\mathrm{SW}(\pi(\hat{\theta})) - \mathrm{SW}(\pi(\hat{\theta}_{-i})) \Big) \\
=& \sum_{i \in N} \mathrm{SW}(\pi(\hat{\theta}_{-i})) - (n-1)\mathrm{SW}(\pi(\hat{\theta})) \\
\geqslant & \sum_{i \in N} \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_i(\hat{\theta})\}) - (n-1)\mathrm{SW}(\pi(\hat{\theta}))
\end{aligned}$$

$$
\begin{aligned}
&\geqslant \mathrm{SW}(\pi(\hat{\theta})) + \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_1(\hat{\theta}), \pi_2(\hat{\theta})\}) + \sum_{i \in N \setminus \{1,2\}} \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_i(\hat{\theta})\}) - (n-1)\mathrm{SW}(\pi(\hat{\theta}))\\
&\geqslant \cdots\\
&\geqslant \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_1(\hat{\theta}), \ldots, \pi_k(\hat{\theta})\}) + \sum_{i \in N \setminus \{1,\ldots,k\}} \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_i(\hat{\theta})\}) - (n-k)\mathrm{SW}(\pi(\hat{\theta}))\\
&\geqslant \cdots \geqslant 0,
\end{aligned}
$$

where the first inequality follows from that $\forall i \in N, (\pi(\hat{\theta}) \setminus \{\pi_i(\hat{\theta})\})$ is one of the allocations given $\hat{\theta}_{-i}$ while $\pi(\hat{\theta}_{-i})$ is an efficient one, and the second inequality follows from $\mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_1(\hat{\theta})\}) + \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_2(\hat{\theta})\}) \geqslant \mathrm{SW}(\pi(\hat{\theta})) + \mathrm{SW}(\pi(\hat{\theta}) \setminus \{\pi_1(\hat{\theta}), \pi_2(\hat{\theta})\})$ given the submodular property of f (as in (1)) as $f(\cup_{j\in N\setminus\{1\}}\pi_j(\hat{\theta})) + f(\cup_{j\in N\setminus\{2\}}\pi_j(\hat{\theta})) \geqslant f(\cup_{j\in N}\pi_j(\hat{\theta})) + f(\cup_{j\in N\setminus\{1,2\}}\pi_j(\hat{\theta}))$. Thus we have proven that this VCG mechanism is profitable.

Lemma 1. Finding an allocation to maximize SW when $|R| = 1$ (i.e., each agent only have one candidate route) is computationally inefficient.

Proof. If $|R| = 1$, any possible allocation π can be seen as a subset of $\cup_{i\in N}\hat{\theta}_i$, i.e., the set of all candidate routes. Even if we do not consider the costs of the observations, for observation value of entropy [31], mutual information [33] and informative path [18], the problem of selecting optimal allocation has been proven NP-hard. Hence, in general, we most likely cannot expect to find the optimal set of locations effectively.

Theorem 4. VCG mechanism is computationally inefficient.

Proof. For problems when $|R| > 1$, there exist special cases in which each agent has only one useful candidate route and $|R|-1$ empty candidate routes. As these cases can be easily transformed to problems of $|R| = 1$, the more general problems of $|R| > 1$ should be computationally harder than the problems of $|R| = 1$. Given above, we have proven that finding an efficient allocation to maximize the social welfare is computationally inefficient and then VCG mechanism is computationally inefficient.

Given Theorem 4, we have known that profit efficiency is incompatible with computational efficiency properties. The VCG mechanism is thus individually rational, profitable, truthful, but it is computationally inefficient. We next design a mechanism that satisfies all of the four desirable properties.

5.2 SSM

In this subsection, we introduce SSM for IPS. As illustrated in Algorithm 2, SSM consists of two phases: the route allocation phase and the bonus determination phase.

5.2.1 *Route allocation*

For any given submodular observation function, we denote the marginal contribution of a route s with respect to observations set Φ by $f_\Phi(s) = f(\Phi \cup s) - f(\Phi)$. We select agents and allocate their routes according to the difference of their marginal contributions and costs, recursively computed by $\pi_i = \arg\max_{s\in S_{-i}} \{f_{\Phi_{i-1}}(s) - c(s)\}$, where $S_{-i} = \{s_{jr}\}_{j\in N\setminus\{1,\ldots,i-1\}, r\in R}$ is the set of all reported routes of by far unallocated agents, $\Phi_{i-1} = \pi_1 \cup \cdots \cup \pi_{i-1}$ is the observation set of by far allocated routes and $\Phi_0 = \emptyset$, and $c(s)$ is the estimated cost of route s. To simplify notations, we will denote this sorted order of agents by N, represent $f_i = f(\Phi_{i-1} \cup \pi_i) - f(\Phi_{i-1})$ and use c_i to denote the cost of π_i. This sorting, in the presence of submodularity, implies that

$$f_1 - c_1 \geqslant f_2 - c_2 \geqslant \cdots \geqslant f_n - c_n. \tag{6}$$

Then the set of allocated routes are $\{\pi_1, \pi_2, \ldots \cup \pi_l\}$, where $l \leqslant n$ is the largest index such that $f_l - c_l > 0$.

5.2.2 *Bonus determination*

In the bonus determination phase, we compute the bonus x_i for each selected agent i.

The intuition behind our bonus schema is as follows. Consider running route allocation without agent i, i.e., sorting the agents in $N \setminus \{i\}$ given their marginal-contribution-minus-cost (as in (6)). For the first j agents in this sorting, we can use the marginal-contributions of candidate routes of agent i at position j to find the maximal advantage that agent i can be allocated instead of the agent at jth place in this

Algorithm 2 The SSM

Input: Report of candidate routes $\hat{\theta} = (\hat{\theta})_{i\in N}$.
Output: A schedule π and bonuses x.
{ Phase 1: Route allocation.}
Initialise: $S \leftarrow \{s_{jr}\}_{j\in N, r\in R}$, $\Phi \leftarrow \emptyset$, $i \leftarrow 1$.
while $S \neq \emptyset$ **do**
  $\pi_i \leftarrow \arg\max_{s\in S} \{f_\Phi(s) - c(s)\}$;
  if $f_i - c_i > 0$ **then**
    $\Phi \leftarrow \Phi \cup \pi_i$;
    $S \leftarrow \{s_{jr}\}_{j\in N\setminus\{1,\ldots,i-1\}, r\in R}$;
  else
    $S \leftarrow \emptyset$;
  end if
  $i \leftarrow i + 1$;
end while
$l \leftarrow i$, $\pi \leftarrow \{\pi_1, \pi_2, \ldots, \pi_l\}$;
{ Phase 2: Bonus determination.}
for $\pi_i \in \pi$ **do**
  $S' \leftarrow \cup_{j\in N\setminus\{i\}}\hat{\theta}_j$, $\Phi' \leftarrow \emptyset$, $\delta^* \leftarrow 0$;
  while $S' \neq \emptyset$ **do**
    $\pi'_{i_j} \leftarrow \arg\max_{s\in S'} \{f_{\Phi'}(s) - c(s)\}$;
    if $f'_{i_j} - c'_{i_j} > 0$ **then**
      for $s_{ir} \in \hat{\theta}_i$ **do**
        $\delta \leftarrow \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right) - (f'_{i_j} - c'_{i_j})$;
        $\delta^* \leftarrow \max\{\delta^*, \delta\}$;
      end for
      $\Phi' \leftarrow \Phi' \cup \pi'_{i_j}$;
      $S' \leftarrow S' \setminus \hat{\theta}_{i_j}$;
    else
      for $s_{ir} \in \hat{\theta}_i$ **do**
        $\delta \leftarrow \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right)$;
        $\delta^* \leftarrow \max\{\delta^*, \delta\}$;
      end for
      $S' \leftarrow \emptyset$;
    end if
  end while
  $x_i \leftarrow \delta^* + c(\pi_i)$;
end for
$x \leftarrow \{x_1, x_2, \ldots, x_l\}$;
return (π, x).

sorting. Then we define the bonus for agent i as the sum of the maximum advantage value and the cost of her allocated route. Although this bonus schema seems a bit tricky, we will later prove that it satisfies several desirable properties.

Specifically, we denote the sorting of the agents in $N' = N \setminus \{i\}$ as follows:

$$f'_{i_1} - c'_{i_1} \geqslant f'_{i_2} - c'_{i_2} \geqslant \cdots \geqslant f'_{i_{|N'|}} - c'_{i_{|N'|}}, \tag{7}$$

where i_j is the jth agent, π'_{i_j} is her allocated route, $f'_{i_j} = f(\Phi'_{j-1} \cup \pi'_{i_j}) - f\left(\Phi'_{j-1}\right)$ denotes the marginal contribution of agent i_j and Φ'_{j-1} denotes the observation set of the first $j-1$ agents' allocated routes. Let l' denote the position of last agent $i_j \in N'$, such that $f'_{i_j} - c'_{i_j} > 0$. For each candidate route $s_{ir} \in \hat{\theta}_i$ of agent i, its marginal contribution at position j is $f_{i(j)}\left(s_{ir}\right) = f\left(\Phi'_{j-1} \cup s_{ir}\right) - f\left(\Phi'_{j-1}\right)$. We then define the advantage value that agent i can be allocated instead of the agent at jth place as

$$\delta_{i(j)}\left(s_{ir}\right) = \begin{cases} \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right) - \left(f'_{i_j} - c'_{i_j}\right), & \text{if } 1 \leqslant j \leqslant l', \\ \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right), & \text{if } j = l'+1, \end{cases}$$

where for $j \in \{1, \ldots, l'\}$, the advantage value is the difference of their marginal-contribution-minus-cost; because no agent in $N \setminus \{i\}$ can be selected at position $l'+1$, the advantage value that to be allocated at position $l'+1$ is set as $\delta_{i(j)}\left(s_{ir}\right) = \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right)$. The intuition of $\delta_{i(j)}$ is the advantage that agent i can be allocated instead of the agent at jth place in this sorting.

Given this, we can get the maximal advantage value for all candidate routes of agent i at all these $l'+1$ points in the sorting over agents in $N \setminus \{i\}$, i.e.,

$$\delta_i^* = \max_{j\in\{1,\ldots,l'+1\}, s_{ir}\in\hat{\theta}_i} \delta_{i(j)}\left(s_{ir}\right). \tag{8}$$

We then set the bonus of agent i to be

$$x_i = \delta_i^* + c(\pi_i). \tag{9}$$

We next analyse the properties of SSM.

Theorem 5. SSM is individually rational.

Proof. In the bonus determination phase, agent i_i is agent i's replacement which appears in the ith place in the sorting of agents in $N \setminus \{i\}$. Because agent i with π_i is preferred than agent i_i at ith place if i is considered, we have that $\delta_i^* \geqslant \left(f_{i(i)}(\pi_i) - c(\pi_i)\right) - \left(f'_{i_i} - c'_{i_i}\right) = (f_i - c_i) - \left(f'_{i_i} - c'_{i_i}\right) \geqslant 0$. If i_i does not exist, it means that i is the last selected agent in the sorting of agents in N, i.e., $i = l$. We then have that $\delta_i^* \geqslant \left(f_{i(i)}(\pi_i) - c(\pi_i)\right) \geqslant 0$. Hence whether i_i is exist or not, we have $x_i = \delta_i^* + c(\pi_i) \geqslant c(\pi_i)$.

Theorem 6. SSM is truthful.

Proof. Given Assumption 1, we only have to prove that an agent cannot get more utility when concealing some candidate routes. Specifically, as a report profile satisfies $\hat{\theta}_i \subseteq \theta_i$, we then have that $\left\{\delta_{i(j)}\left(s_{ir}\right)\right\}_{s_{ir}\in\hat{\theta}_i} \subseteq \left\{\delta_{i(j)}\left(s_{ir}\right)\right\}_{s_{ir}\in\theta_i}$. From (8), we then have that the maximal advantage value cannot increase if less candidate routes are reported. From (4) and (9), we can easily get that an agent cannot get more utility by reporting a false profile.

Theorem 7. SSM is profitable.

Proof. Because the utility of the platform is $u(\pi, x) = f(\cup_{i\in N}\pi_i) - \sum_{i\in N} x_i = \sum_{i\in\{1,\ldots,l\}} f_i - \sum_{i\in\{1,\ldots,l\}} x_i$, it suffices to prove that $f_i > x_i$ for each $i \in \{1, \ldots, l\}$. Recall that there are l' agents in the sorting in (7) when determining the bonus of agent i, such that $f'_{i_j} > c'_{i_j}$.

For $j \leqslant i$, because the fact that $\pi'_{i_j} = \pi_j$ is preferred than all of i's candidate routes in the sorting in (6), we have that $\Phi'_{j-1} = \Phi_{j-1}$ and $f'_{i_j} - c'_{i_j} = f_j - c_j$. We then have that $\forall s_{ir} \in \hat{\theta}_i, \delta_{i(j)}(s_{ir}) = \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right) - \left(f'_{i_j} - c'_{i_j}\right) = \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right) - (f_j - c_j) \leqslant 0$. Moreover, we have that $f_i > c_i$ as agent i is selected in the route allocation phase. We then have that $\forall j \leqslant i$ and $s_{ir} \in \hat{\theta}_i$, $f_i - c_i > 0 \geqslant \delta_{i(j)}(s_{ir})$.

For $i < j \leqslant l' + 1$ and $s_{ir} \in \hat{\theta}_i$, as $f\left(\Phi'_{i(i)} \cup (s_{ir})\right) - f\left(\Phi'_{i(i)}\right) \geqslant f\left(\Phi'_{i(j)} \cup (s_{ir})\right) - f\left(\Phi'_{i(j)}\right)$ given the decreasing marginal value property of f and $\Phi'_{i(i)} \subseteq \Phi'_{i(j)}$, we have that $f_{i(i)}(s_{ir}) - c(s_{ir}) \geqslant f_{i(j)}(s_{ir}) - c(s_{ir}) \geqslant \min\{f_{i(j)}(s_{ir}) - c(s_{ir}), \left(f_{i(j)}(s_{ir}) - c(s_{ir})\right) - \left(f'_{i_j} - c'_{i_j}\right)\} = \delta_{i(j)}\left(s_{ir}\right)$. We then have that $\forall s_{ir} \in \hat{\theta}_i, f_i - c_i = f_{i(i)}(\pi_i) - c(\pi_i) \geqslant f_{i(i)}(s_{ir}) - c(s_{ir}) \geqslant \delta_{i(j)}\left(s_{ir}\right)$ when $i < j \leqslant l' + 1$.

Above we have proven that $\forall j \in \{1, \ldots, l' + 1\}$ and $s_{ir} \in \hat{\theta}_i, f_i - c_i \geqslant \delta_{i(j)}(s_{ir})$. Thus we have $f_i - c_i \geqslant \delta_i^*$, i.e., $f_i \geqslant (\delta_i^* + c_i) = x_i$.

Theorem 8. SSM is computationally efficient.

Proof. In the route allocation phase, finding a route with maximal marginal value (line 5) takes $O\left(|N| \cdot |R|\right)$ time, where $|N| \cdot |R|$ is the number of all candidate routes. Since the number of selected agents is at most N, the while-loop (lines 4–12) thus takes $O(|N|^2 \cdot |R|)$. In the bonus determination phase, for each allocated agent, it takes $O\left(|N| \cdot |R| + |R|\right)$ time to find the maximal advantage value by running the while-loop (lines 17–33). Hence, the for-loop (lines 15–35) takes $O\left(|N|^3 \cdot |R|^2\right)$ time. Hence, our mechanism takes $O\left(|N|^3 \cdot |R|^2\right)$ time, which means that it is polynomial time computable.

Here we use an example to show that our SSM is helpful. We assume that there are two participants and each has two candidate routes. The observation values and costs of these routes are as follows:

$$\begin{aligned}
&f(s_{11}) = 6, \quad f(s_{12}) = 5, \quad f(s_{21}) = 4, \quad f(s_{22}) = 3,\\
&f(s_{11} \cup s_{21}) = 7, \quad f(s_{11} \cup s_{22}) = 8, \quad f(s_{12} \cup s_{21}) = 9, \quad f(s_{12} \cup s_{22}) = 6,\\
&c(s_{11}) = c(s_{12}) = c(s_{21}) = c(s_{22}) = 1.
\end{aligned}$$

Using SSM, we allocate the routes for the agents by the marginal contributions of their routes.

First round: $f(s_{11}) - c(s_{11}) > f(s_{12}) - c(s_{12}) > f(s_{21}) - c(s_{21}) > f(s_{22}) - c(s_{22})$, then s_{11} is selected.
Second round: $f(s_{11} \cup s_{22}) - f(s_{11}) - c(s_{22}) > f(s_{11} \cup s_{21}) - f(s_{11}) - c(s_{21})$, then s_{22} is selected.
Given above, $\{s_{11}, s_{22}\}$ is allocated.

Then the bonuses of the two agents is calculated as

x_1 : $\{s_{21}\}$ is selected when ranking the routes without agent 1.
$\delta_{1(1)}(s_{11}) = (f_{1(1)}(s_{11}) - c(s_{11})) - (f(s_{21}) - c(s_{21})) = 2, \delta_{1(2)}(s_{11}) = (f_{1(2)}(s_{11}) - c(s_{11})) = 2,$
$\delta_{1(1)}(s_{12}) = (f_{1(1)}(s_{12}) - c(s_{12})) - (f(s_{21}) - c(s_{21})) = 1, \delta_{1(2)}(s_{12}) = (f_{1(2)}(s_{12}) - c(s_{12})) = 4,$
$x_1 = \max\{2, 2, 1, 4\} + c(s_{11}) = 5.$
x_2 : $\{s_{11}\}$ is selected when ranking the routes without agent 2.
$\delta_{2(1)}(s_{21}) = (f_{2(1)}(s_{21}) - c(s_{21})) - (f(s_{11}) - c(s_{11})) = -2, \delta_{2(2)}(s_{21}) = (f_{2(2)}(s_{21}) - c(s_{21})) = -3,$
$\delta_{2(1)}(s_{22}) = (f_{2(1)}(s_{22}) - c(s_{22})) - (f(s_{11}) - c(s_{11})) = -1, \delta_{2(2)}(s_{22}) = (f_{2(2)}(s_{22}) - c(s_{22})) = 1,$
$x_2 = \max\{-2, -3, -1, 1\} + c(s_{22}) = 2.$

As we can see in this example, although s_{11} is allocated to agent 1, s_{12} helps her get a higher bonus (and utility). In other words, SSM can incentive agent 1 to report all of her two candidate routes. In addition, if we extend our model to allow each agent to have its own cost function and can strategically report the cost value of a candidate route, agent 1 in the this example may report a lower $c(s_{12})$ (say 0.5) to get a higher bonus (5.5) when running SSM, which means that SSM will not be truthful.

Thus, we have theoretically proven that SSM is individually rational, profitable, truthful and computationally efficient for our IPS model. We further design experiments to evaluate SSM's performance in the next section.

6 Performance evaluation

To evaluate SSM, we design several benchmark mechanisms, then set up the experiment and present the results and discussion. The performance metrics include social welfare, platform's utility and user's utility in general.

6.1 Benchmarks

We compare our SSM with four benchmark mechanisms: (1/3)-Approximation, Local-Best, Random, and Upper-Bound. First, the route allocation algorithms of these mechanisms are as follows.

(1/3)-Approximation. Its allocation algorithm relies on a local-search technique, which sequentially searches for a better solution by adding a new candidate route or deleting an existing route whenever possible. For special cases where $|R| = 1$, i.e., each agent have only one candidate route, the work in [34] proved that this will be a deterministic linear time algorithm and (1/3) approximation for the problem of maximising this type of non-monotone submodular functions. However, finding an approximation algorithm for the IPS cases where $|R| > 1$ is still an open problem, and designing a truthful payment method for this local-search allocation approach is also an open problem.

Local-Best. Its allocation algorithm assumes that each agent selects her best candidate route by ignoring other agents' participation. It is an algorithm that creates a policy by agents making local decisions, i.e., without coordination, in environmental monitoring campaigns.

Random. Its allocation algorithm assumes that routes are randomly selected by agents. Similar to Local-Best, Random is also an algorithm in which agents make local decisions.

Also, because the optimal route allocation algorithm is computationally infeasible, we developed an Upper-Bound to the algorithm that can be easily calculated. The Upper-Bound is described below.

Upper-Bound. We relax the assumption that each person can take observations along only one route at the same time and we assume that there are no costs for people taking these observations. Thus, all participants are assumed to take measurements at every timestep of all candidate routes and the total utility can be trivially calculated.

Next, we simply designed a bonus determination method for these mechanisms as follows. If an agent is assigned a route to her, then her bonus is set to be equal to the cost of the route. Otherwise, if no route is assigned, the agent will get nothing. Although every agent gets zero utilities with this method, these mechanisms will be individual rational and truthful. The profitable property of these mechanisms can not be guaranteed.

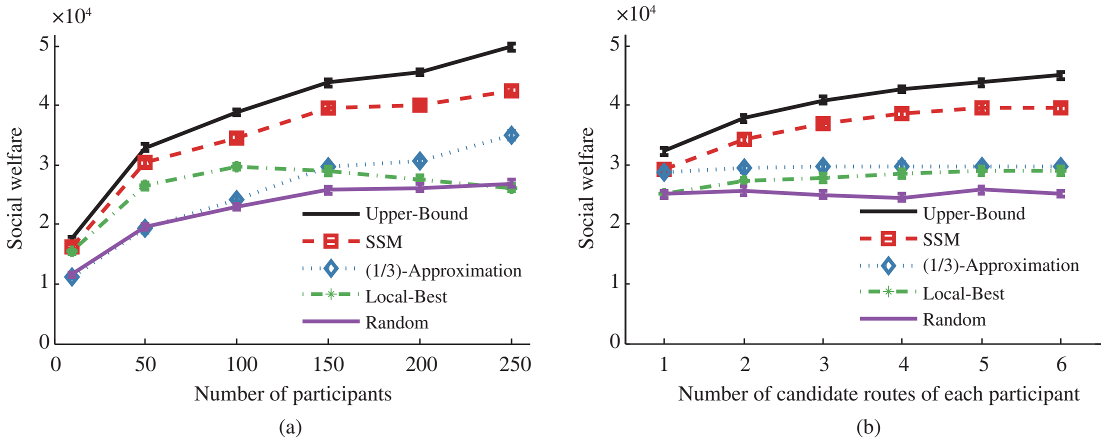

Figure 3 (Color online) Average social welfare for (a) different number of participants and (b) different number of each participant's candidate routes.

6.2 Experimental setup

Our experiments are based on the setup described in [14, 16]. In particular, we focus on air quality in terms of fine particular matter (PM 2.5) in Beijing, where people are plagued by winter air pollution. We set up the experiments by using GP to model the value of information about this environmental phenomenon and using a mobility pattern prediction system with historic mobility data to produce a set of candidate routes for each participant. Given the historic fine-grained data provided from a number of static air quality stations in Beijing [35], we use a common technique called maximum likelihood estimation (MLE) to train the GP model in our experiments.

To extract participants' routine routes, we use the data from Geolife trajectories dataset[1] which contains sequences of time-stamped locations of 182 people in Beijing over a period of 5 years (2007–2012). We take the locations of 108 people each hour over a period for our experiments and randomly take several sets of consecutive observations from each person's data as her candidate routes. To test our system for more than 108 participants, we take the data of different months but the same days from the same pool of people's trajectories. The cost of taking measurements along a route is randomly assigned to each person and we assume that taking measurements at peak hours is more expensive.

Moreover, in order to obtain statistical significant in our results, we perform two-sided t-test significance testing at the 95% confidence interval. The algorithms run on a machine with 2.6 GHz Intel dual core CPU and 1 GB RAM.

6.3 Results and discussion

Here we describe the results of three experiments: (1) we simulate a varying number of participants with 5 candidate routes for each participant; (2) we simulate 150 participants with a varying number of each participant's candidate routes; (3) we simulate a varying average cost of candidate routes with 150 participants and 250 participants, respectively.

First, as we can see in Figures 3 and 4, the average social welfare earned in these experiments by SSM is consistently higher than that of (1/3)-Approximation, Local-Best and Random. For more than 50 participants, SSM outperforms both (1/3)-Approximation, Local-Best and Random by more than 15%. Surprisingly, SSM reaches more than 86% of Upper-Bound. As shown in Figure 3(a), for more than 150 participants, the average social welfare earned by Local-Best is decreasing with the number of participants, which means that the costs of hiring more participants could be higher than the marginal information they obtain by using randomly allocated routes. Then, as shown in Figure 3(b), with a varying number of each participant's candidate routes, Upper-Bound and SSM obtain more social welfare when people have more candidate routes for the platform to allocate, and the social warfare earned by (1/3)-Approximation, Local-Best and Random are not much different. Moreover, as we can see in Figure 4, the obtained social welfare by Local-Best and Random dramatically decreases with the average cost of candidate routes while

1) https://www.microsoft.com/en-us/research/publication/geolife-gps-trajectory-dataset-user-guide/.

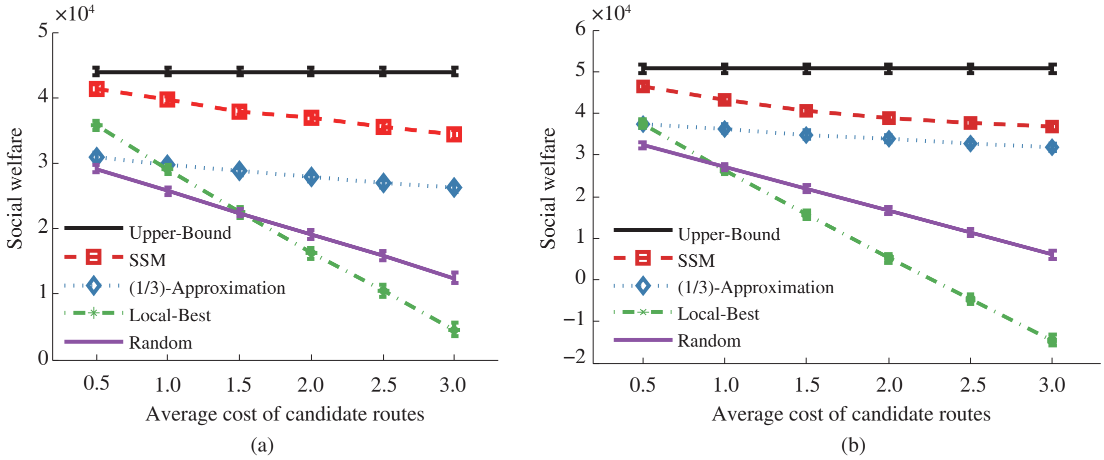

Figure 4 (Color online) Average social welfare for different average cost of candidate routes with (a) 150 agents and (b) 250 agents.

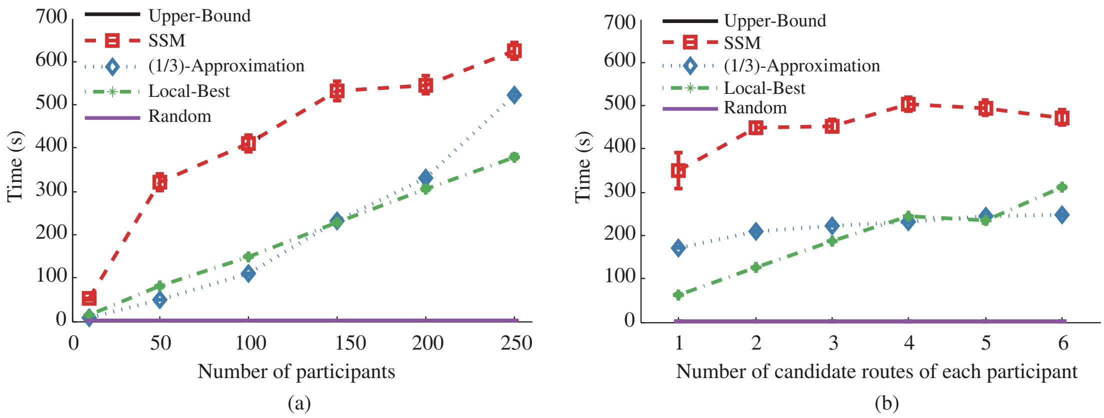

Figure 5 (Color online) Average runtime for (a) different numbers of participants and (b) different numbers of each participant's candidate routes.

SSM and (1/3)-Approximation achieve acceptable performance owing to their coordination methods. As our Upper-Bound assumes that there are no costs for people taking observations, its social welfare keeps the same value with different average costs.

Second, Figure 5(a) and (b) shows the performance of the algorithms in terms of the average runtime when varying the number of participants and each participant's candidate routes. The runtimes of SSM, (1/3)-Approximation and Local-Best are increasing with the number of participants obviously, respectively. However, SSM's computation time is not always growing with the number of each participant's candidate routes. For more than 3 candidates routes, the computation time is decreasing. Moreover, although SSM's computation time is much longer than other algorithms, its growth rate is acceptable.

7 Conclusion

In this paper, we proposed a new model of self-interested users engaging in IPS. In particular, we considered the situations where the service provider can coordinate the participants' schedules to maximize the information collected while mitigating the costs incurred. We proposed a VCG based mechanism for our IPS problem and proved that it is individually rational, truthful, profitable, but computationally inefficient. Moreover, we proposed a sequentially sorting based mechanism, SSM, and proved that it is individually rational, truthful, profitable, and computationally efficient. Then we empirically showed that the allocation algorithm of SSM outperforms two baseline algorithms by more than 15%. Besides the proposed four desirable properties, future work will consider designing a mechanism with an approximation optimal allocation method. In addition, future work will also consider more complex situations in which both candidate routes and their costs are private and can be strategically reported. The truthful feature will not be guaranteed if we directly use the route allocation and the bonus determination method of our SSM in these situations, so new mechanisms will be required.

Acknowledgements This work was supported by National Natural Science Foundation of China (Grant Nos. 61702528, 61603403, 61603406).

References

1 Burke J, Estrin D, Hansen M, et al. Participatory sensing. In: Proceedings of Workshop on World-Sensor-Web (WSW06): Mobile Device Centric Sensor Networks and Applications, 2006
2 Waze, https://www.waze.com/
3 Weathermob, http://weathermob.me/
4 Kanjo E. NoiseSPY: a real-time mobile phone platform for urban noise monitoring and mapping. Mobile Netw Appl, 2010, 15: 562–574
5 Ushahidi, https://www.ushahidi.com/
6 DHondt E, Stevens M, Jacobs A. Participatory noise mapping works! An evaluation of participatory sensing as an alternative to standard techniques for environmental monitoring. Pervasive Mobile Comput, 2013, 9: 681–694
7 Pan Z, Yu H, Miao C, et al. Crowdsensing air quality with camera-enabled mobile devices. In: Proceedings of the 21st AAAI Conference on Artificial Intelligence, San Francisco, 2017. 4728–4733
8 Radanovic G, Faltings B. Incentive schemes for participatory sensing. In: Proceedings of 2015 International Conference on Autonomous Agents and Multiagent Systems, International Foundation for Autonomous Agents and Multiagent Systems, 2015. 1081–1089
9 Yang D, Xue G, Fang X, et al. Incentive mechanisms for crowdsensing: crowdsourcing with smartphones. IEEE/ACM Trans Network, 2016, 24: 1732–1744
10 Chen H H, Guo B, Yu Z W, et al. Mobile crowd photographing: another way to watch our world. Sci China Inf Sci, 2016, 59: 083101
11 Ma X D, Li H, Ma J F, et al. APPLET: a privacy-preserving framework for location-aware recommender system. Sci China Inf Sci, 2017, 60: 092101
12 Singla A, Krause A. Incentives for privacy tradeoff in community sensing. In: Proceedings of the 1st AAAI Conference on Human Computation and Crowdsourcing, 2013
13 Restuccia F, Ghosh N, Bhattacharjee S, et al. Quality of information in mobile crowdsensing: survey and research challenges. ACM Trans Sen Netw, 2017, 13: 1–43
14 Zenonos A, Stein S, Jennings N R. Coordinating measurements for air pollution monitoring in participatory sensing settings. In: Proceedings of the 14th International Conference on Autonomous Agents and Multiagent Systems, Istanbul, 2015. 493–501
15 Zenonos A, Stein S, Jennings N R. An algorithm to coordinate measurements using stochastic human mobility patterns in large-scale participatory sensing settings. In: Proceedings the 30th AAAI Conference on Artificial Intelligence, 2016. 3936–3943
16 Zenonos A, Stein S, Jennings N R. Coordinating measurements in uncertain participatory sensing settings. jair, 2018, 61: 433–474
17 Groves T. Incentives in teams. Econometrica, 1973, 41: 617
18 Singh A, Krause A, Kaiser W J. Nonmyopic adaptive informative path planning for multiple robots. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, 2009. 1843–1850
19 Lane N, Miluzzo E, Lu H, et al. A survey of mobile phone sensing. IEEE Commun Mag, 2010, 48: 140–150
20 Khan W Z, Xiang Y, Aalsalem M Y, et al. Mobile phone sensing systems: a survey. IEEE Commun Surv Tut, 2013, 15: 402–427
21 Ben-Zeev D, Scherer E A, Wang R, et al. Next-generation psychiatric assessment: using smartphone sensors to monitor behavior and mental health. Psychiat Rehabil J, 2015, 38: 218–226
22 Gao H, Liu C H, Wang W, et al. A survey of incentive mechanisms for participatory sensing. IEEE Commun Surv Tut, 2015, 17: 918–943
23 Singer Y. Budget feasible mechanisms. In: Proceedings of 2010 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2010. 765–774
24 Singh A, Krause A, Guestrin C, et al. Efficient informative sensing using multiple robots. J Artifi Intell Res, 2009, 34: 707–755
25 Stranders R, Farinelli A, Rogers A, et al. Decentralised coordination of mobile sensors using the max-sum algorithm. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, 2009. 299–304
26 Stranders R, Munoz de Cote E, Rogers A, et al. Near-optimal continuous patrolling with teams of mobile information gathering agents. Artifi Intell, 2013, 195: 63–105
27 Rasmussen C E, Williams C K I. Gaussian Processes for Machine Learning. Cambridge: MIT Press, 2006

28 Krause A, Guestrin C. Submodularity and its applications in optimized information gathering. ACM Trans Intell Syst Technol, 2011, 2: 1–20
29 Zenonos A, Stein S, Jennings N. A trust-based coordination system for participatory sensing applications. In: Proceedings of the 5th AAAI Conference on Human Computation and Crowdsourcing, 2017
30 Guestrin C, Krause A, Singh A P. Near-optimal sensor placements in gaussian processes. In: Proceedings of the 22nd International Conference on Machine Learning, 2005. 265–272
31 Ko C W, Lee J, Queyranne M. An exact algorithm for maximum entropy sampling. Operations Res, 1995, 43: 684–691
32 Vickrey W. Counterspeculation, auctions, and competitive sealed tenders. J Finance, 1961, 16: 8–37
33 Krause A, Singh A, Guestrin C. Near-optimal sensor placements in gaussian processes: theory, efficient algorithms and empirical studies. J Mach Learn Res, 2008, 9: 235–284
34 Feige U, Mirrokni V S, Vondrák J. Maximizing non-monotone submodular functions. SIAM J Comput, 2011, 40: 1133–1153
35 Zheng Y, Liu F, Hsieh H-P. U-air: when urban air quality inference meets big data. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'13, New York, 2013. 1436–1444