

# Optimal Incentive-driven Design of Participatory Sensing Systems

Iordanis Koutsopoulos  
University of Thessaly and  
Centre for Research and Technology Hellas (CERTH), Greece

**Abstract**—Participatory sensing has emerged as a novel paradigm for data collection and collective knowledge formation about a state or condition of interest, sometimes linked to a geographic area. In this paper, we address the problem of incentive mechanism design for data contributors for participatory sensing applications. The service provider receives service queries in an area from service requesters and initiates an auction for user participation. Upon request, each user reports its perceived cost per unit of amount of participation, which essentially maps to a requested amount of compensation for participation. The participation cost quantifies the dissatisfaction caused to user due to participation. This cost is considered to be private information for each device, as it strongly depends on various factors inherent to it, such as the energy cost for sensing, data processing and transmission to the closest point of wireless access, the residual battery level, the number of concurrent jobs at the device processor, the required bandwidth to transmit data and the related charges of the mobile network operator, or even the user discomfort due to manual effort to submit data. Hence, participants have strong motive to misreport their cost, i.e. declare a higher cost than the actual one, so as to obtain higher payment.

We seek a mechanism for user participation level determination and payment allocation which is most viable for the provider, that is, it minimizes the total cost of compensating participants, while delivering a certain quality of experience to service requesters. We cast the problem in the context of optimal reverse auction design, and we show how the different quality of submitted information by participants can be tracked by the service provider and used in the participation level and payment selection procedures. We derive a mechanism that optimally solves the problem above, and at the same time it is individually rational (i.e., it motivates users to participate) and incentive-compatible (i.e. it motivates truthful cost reporting by participants). Finally, a representative participatory sensing case study involving parameter estimation is presented, which exemplifies the incentive mechanism above.

## I. INTRODUCTION

In recent years, participatory (or community) sensing has rapidly proliferated as a paradigm for multi-modal data collection and dissemination [1], [2]. The large penetration of smartphones with various embedded sensors (e.g. camera, microphone, accelerometer, light, GPS) have greatly automatized the information generation process, which now mostly occurs without human intervention. On the other hand, in cases where human intervention is needed, the ease with which information like text, image or video can be created and uploaded, has made the human factor almost transparent in the whole chain of information generation, transport and consumption. Participatory sensing has also revolutionized the field of wireless sensor networks, since it eliminates the need for deploying a specific purpose sensor network. It has also created a novel

viewpoint in information gathering and exploitation, which is known as *collective awareness*, or *collective intelligence*. That is, information can be easily collected through user contribution, and after appropriate processing and aggregation, it can aid in forming collective knowledge about a specific state or condition of interest, sometimes linked to a geographic area. Collective knowledge is of value to the community at large, and, depending on the application, it can be directly or indirectly influential to data contributors as well.

In participatory sensing, a specific-purpose application is launched by an application service provider. Applications that rely on participatory sensing are abundant and diverse, and they range from air pollution or electromagnetic field (EMF) radiation monitoring, up to road traffic condition reporting, prediction and tracking of disease outbreaks, or urban parking space management. The common objective in all applications is to represent an underlying phenomenon, process or state as accurately as possible and deliver it to interested users. Thus, in air pollution and EMF radiation monitoring, the target is a dynamic city map of air pollution or EMF radiation levels; in parking space management, the goal is to construct a map with free parking spot locations, and so on.

There exist participants to the service, who act as information providers. These entities submit measurement data, usually through mobile devices, or they simply provide input, whatever that might be, based on their own experience at their location. Data are submitted through the wireless operator infrastructure or WiFi access points and are forwarded to the application service provider. There also exist subscribers to the service, namely information consumers. These place queries related to the service (e.g. about the air pollution level or free parking spots) in their vicinity and require to be served. The service provider aggregates data samples of providers and forwards the result to querying entities for some price, thereby generating revenue. However, service availability and provisioning at certain quality depends crucially on *participation levels of providers* and on the *quality* of provided information. Information providers voluntarily contribute data subject to various costs, such as mobile device battery energy cost for sensing, data processing and transmission, charges by the mobile operator for the bandwidth needed for transmitting data, processing power cost, or discomfort due to manual effort to submit data. It is therefore very important for the survivability of the service to have appropriate *incentive mechanisms* to motivate users to participate in data collection.

Incentive mechanisms can be actual payments or credits. Payments need to be large enough to cover participation costs and create motives for participation but also low enough, so that the application service is maintained without much expenditure by the service provider. The participation cost is private information for users, and users are strongly motivated to misreport the actual cost so as to receive higher compensation. The design of incentive mechanisms for participatory sensing applications that are economically viable for the application provider, yet they encourage *truthful* cost reporting for users and guarantee a given quality of service, is the problem we address in this paper, by using the *optimal auction* design framework [18].

#### A. Related work

Participatory sensing applications can be classified into three categories: environment-centered, infrastructure and facility related, and socially or community centered. In the first class, the OpenSense [3] participatory sensing infrastructure performs real-time air quality monitoring and comprises heterogeneous sensors and a management middleware. In the second category, an example is GreenGPS [4], which uses a vehicle interface to measure and transmit fuel consumption and location data. It then constructs fuel-efficient routes to destinations for querying users. CrowdPark [5] facilitates parking reservation through user submitted information about when parking resources will be available, and uses this information to help other users locate parking spots. In the same spirit but using a different approach, ParkNet [6] application detects available parking spots using ultrasonic sensing devices installed on cars, combined with smart phones. Finally, in a representative system of the third category, LiveCompare participants use their phone cameras to take a picture of the price tag of a product of interest [7]. In exchange for submitting a price data point, the user receives pricing information for the product at nearby grocery stores. Finally, in DietSense [8], individuals take pictures of what they eat and share it within a community to compare eating habits, e.g. in a community of diabetics.

In [9], the authors address the problem of sensor selection out of a set of available ones, such that value of derived measurement data is maximized, subject to sensor resource usage constraints. The value of information is defined as the expected weighted reduction in prediction uncertainty of the underlying process at unobserved locations, where the weight is the query demand load at various locations, and the expectation is taken over the probability distribution induced on possible available sensor locations by the selected ones.

The problem of incentive provisioning to users for contributing data is a central one in participatory sensing [10]. In some cases, participation incentives are intrinsic to the application: for example in LiveCompare, users obtain the service, namely product prices in nearby stores, only if they participate by submitting prices themselves. If participants do not obtain a direct benefit from participation, appropriate incentive mechanisms need to be designed. In the CrowdPark system, compensation comes through subsequent gains from re-selling the parking spot, which exceed the refund received

if they deny the spot on purpose. The work in [11] guides credit allocation to participants by using the feedback of quality of provisioned information (QoI) and credit satisfaction from information requesters and participants respectively. In [12], the authors consider users as data contributors and requesters at the same time, and they formulate convex optimization problems that allocate credit quota to users to maximize utility functions that denote social welfare or fairness. User utility is a concave function of normalized allocated credit, normalized over user demand and contribution cost. In [13], the authors consider a reputation-based scheme in the context of a crowdsourcing website. The scheme builds upon a reputation metric that rewards or penalizes users, depending on whether their strategy is aligned to or deviates from a social norm. The scheme is then incorporated into a repeated game model that captures interactions of users who contribute to and request information from the website.

In Bayesian games [14], information about player characteristics (e.g. payoffs) is incomplete or uncertain. Uncertainty is captured by a random variable, the user *type*, whose realization is known only to that user. Mechanism design is the branch of game theory that seeks to influence the outcome of a Bayesian game towards a certain objective. Auctions are an important class of mechanism design and study rules for allocation of a divisible or indivisible good to interested buyers and subsequent pricing [15], under the regime of unknown private buyer valuations. The Vickrey Clark Groves (VCG) auction [16], [17] for divisible goods is efficient, in the sense that it maximizes social welfare of the allocation while ensuring truthful declaration of the private utility function. On the other hand, an *optimal auction* maximizes the expected revenue of the seller while ensuring that truth-telling is an equilibrium of the Bayesian game among users. The seminal work in optimal auctions for an indivisible good is [18].

Along that spirit, the recent work in [19] considers the selection of the subset of users to collect measurements from for maximizing system utility minus sum of payments. The authors exploit sub-modularity of the objective above to show that a greedy algorithm in user selection and payment solves the problem, while the mechanism is truthful and has positive objective function value. In [20], a reverse auction is proposed, in which users submit their offers, and the provider selects some of them and compensates them based on their offers. A virtual credit is also included just for participating in the auction, provided that the participant reduces its offer in subsequent rounds, so that the set of winning users changes at different times. This discourages increasing offers by winning users and is shown to reduce compensation cost. Reverse auctions have also been considered for designing incentive mechanisms for 3G traffic offloading [21], where users are compensated based on the bids they submit about the delay they wish to experience while offloading, versus price reduction.

#### B. Our contribution

We address the problem of designing incentive mechanisms for data contributors for participatory sensing applications

based on a *reverse* auction. The service provider receives service queries in an area and initiates a reverse auction for soliciting user participation. Users may participate in different capacities, e.g. by submitting different number of data samples or different types of data (e.g. text, photo, video, etc). The amount of user participation through data contribution is abstracted as the user *participation level*. Upon request, each interested user reports its incurred cost per unit of amount of participation, which essentially maps to a requested amount of compensation in order to participate. The service provider determines the participation level and payment and announces them to the users. Users submit their data and get reimbursed.

There are several challenges in this setup. First, *the mechanism should motivate users to participate*. That is, each user should obtain utility at least as much as that obtained by not participating in the process. Second, *the service provider does not know the actual participation cost as perceived by users*. This can capture various types of costs such as the energy cost for sensing, processing and data transmission, the required bandwidth to transmit data, the processing power cost, or the discomfort due to manual effort to submit data. The actual participation cost in general captures dissatisfaction due to participation and is private information for each user, since it strongly depends on various factors inherent to the device. For example, the perceived energy cost depends on the proximity of the mobile device to an infrastructure base station or access point, and on the residual amount of device battery energy. The transmit bandwidth cost depends on the associated charges by the mobile operator, or on whether other bandwidth-consuming actions are carried out by the device at the same time. Similarly, the processing power cost depends on the state of concurrent job processing at the device processor. The cost is higher if there exist several computationally-intensive jobs executed at the processor at that time. Participants have strong motive to *misreport* their cost, i.e. declare a higher cost than the actual one, so as to obtain higher payment. It is therefore important for the mechanism to induce users to declare the actual cost.

Third, the decision on user participation levels should definitely consider the *quality of information* that each user provides. For example, if a user consistently provides low quality samples, this user should be given little or no participation and should be compensated little. Fourth, the derived user participation levels should lead to given *guaranteed quality of service* for querying users, otherwise the service provider clientele will be dissatisfied. At the same time, payments to participants should be such that the service as a whole is *viable* for the provider, namely the total expenditure of the provider in order to support the service should be minimized.

The contribution of this work to the literature is as follows:

- We formulate the problem above from the point of view of the service provider that aims at minimizing the cost of compensating participants, while delivering a given quality of experience to service requesters; we cast the problem in the context of *optimal reverse auction*.
- We capture the absence of knowledge of the service

provider about user participation cost, which is private information for each user and is handled through a Bayesian game among users.

- We consider the different quality of submitted information by participants, and we show how the service provider can keep track of it and use it to guide the participation level and payment allocation.
- We derive a policy for user participation level and payment allocation that solves the problem above, and it is individually rational (i.e. it motivates users to participate) and incentive-compatible (i.e. it motivates truthful cost reporting by participants).
- We present a representative case study on parameter estimation, which exemplifies the incentive mechanism above.

Our work is different from the mechanism in [18] in that we consider an optimal *reverse* auction with *multiple* winners. The provider needs to fulfil a given quality of service constraint by “buying” different participation levels from different users. Differently from [20] and [11], in addition to the above, the constraint of providing a guaranteed quality of service and the consideration of quality of contributed data are novel and are taken into account in the allocation process. Moreover, we explicitly compute the payment and participation level for each user. Compared to [19], our method explicitly minimizes compensation cost, and it takes into account the computed quality of submitted information. Furthermore, it adheres to a participation level allocation approach rather than a user selection one, and thus it allows for more flexibility.

The rest of the paper is organized as follows. In section II we present the model and assumptions of our approach. In section III we formulate the problem, we derive the optimal auction design framework and prove the incentive-compatibility and individual rationality of the method. Section IV presents a representative case study, and section V concludes our study.

## II. SYSTEM MODEL

### A. Participants to the service

A participatory sensing application provider launches a specific-purpose application. At a given geographical area, service subscribers place queries which need to be satisfied. In response to queries, the provider broadcasts a request for data contribution in that area. A set  $\mathcal{N}$  of  $N$  user devices that exist in the area respond to the request. These devices are potential *participants* that submit data and get compensated. Let  $p_i$  denote the payment from the service provider to user  $i$ . The main components of the system are depicted in Fig. 1.

1) *Participation level*: The extent to which user  $i \in \mathcal{N}$  contributes by submitting data is quantified by a real-valued variable  $x_i \geq 0$  that denotes user participation level. For participatory sensing applications that involve continuous monitoring of a quantity of interest (e.g. EMF radiation or air pollution) or in general rely on continuous measurements provided by mobile devices (e.g. velocity, acceleration, so as to estimate road traffic conditions), the user participation level is quantified as the number of measurement samples per unit of time. Thus,  $x_i$  is

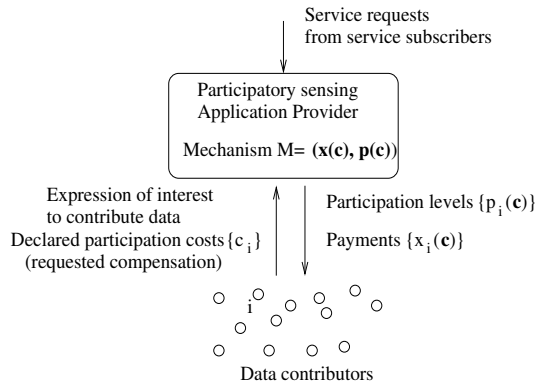


Fig. 1. The main components of the participatory sensing system. The application provider receives service requests from subscribers and sends a request for data contribution in the area. Each interested device  $i$  responds with a declared participation cost  $c_i$  (which maps to a requested amount of compensation per unit of participation level). The provider determines the allocated participation level  $x_i(c)$  and payment  $p_i(c)$  for each device  $i$ , where  $c$  is the vector of declared costs.

the average sampling *rate*, i.e. the average rate with which the device submits measurements to the provider. This definition is generic enough and also captures other types of participatory sensing applications in which users contribute by submitting their experience in terms of an opinion, a picture, and so on. In these cases, sampling rates refer to the specific data points (opinions, pictures, etc) that users submit.

2) *Participation cost*: Each user is characterized by a perceived cost  $C_i > 0$  per unit of participation level. The cost  $C_i$  can also be considered as user  $i$ 's minimum requested compensation per unit of participation level. The actual participation cost is *private* information for each user, as it depends on various factors inherent to the device and the user. For example, the perceived energy cost depends on proximity of the mobile device to a point of wireless access and on the level of residual battery energy. The transmit bandwidth cost is related to the charges of the mobile operator, or on whether other bandwidth-consuming actions are carried out. The processing power cost depends on the state of the number of concurrent jobs processed by the device processor. When it comes to manual data insertion by the user, several subjective factors arise as well. In conclusion, the cost quantifies the dissatisfaction caused to user  $i$  due to participation to the service. Dissatisfaction can be expressed in terms of resource consumption of the device (e.g. energy, computational power, communication bandwidth) or time, attention and effort that needs to be put by the user.

It makes sense to model the cost  $C_i$  for user  $i$  as a continuous random variable that takes values in  $C_i = [\underline{c}_i, \bar{c}_i]$ . Random variables  $C_i$  are independent from each other. Since  $C_i$  is private information for each user  $i$ , only  $i$  knows its realization,  $c_i$ , through superposition of the various factors above that influence the cost. The cost realization  $c_i$  is the user *type* in the Bayesian game setting. On the other hand, the service provider and the users other than  $i$ , have only probabilistic knowledge of  $C_i$ , which may capture prior information about the cost.

Let  $f_i(\cdot)$  denote the probability density function (p.d.f) of  $C_i$ , and let  $F_i(\cdot)$  be the corresponding cumulative density function (c.d.f). The p.d.f and the lower and upper limits,  $\underline{c}_i, \bar{c}_i$ , of its support set could be formed for example from the empirical distribution out of prior cost declarations by the user. The absence of prior information can be captured by taking  $f_i(\cdot)$  to be uniform over  $C_i$ . The utility of user  $i$  for participation level  $x_i$  and payment  $p_i$  is given by

$$u_i = p_i - C_i x_i. \quad (1)$$

### B. Quality of data and quality of service

Contributed data points are received and aggregated by the service provider, depending on the application. For example, in air pollution monitoring, the provider determines the air pollution level through a sufficient statistic of contributed data, e.g. averaging or a linear combination. The quality of service to subscribers depends on (i) the participation level  $x_i$  of each user  $i$ , (ii) the quality of submitted data by each user.

Now, the latter cannot be determined at the time data is submitted. Thus, the provider needs to rely on past experience for that. For each user  $i$ , the provider maintains and continuously updates an empirical *quality indicator*  $q_i$ , which essentially measures the relevance or usefulness of information provided by user  $i$  in the past. This can be quantified by the average deviation of submitted samples from the result of the aggregation of all user samples. For instance, for the class of applications of continuous monitoring of air pollution level  $\theta$ , the provider collects user measurements and aggregates them in some manner to compute an estimate  $\hat{\theta}_t$  of pollution level at time  $t$ . The quality indicator for user  $i$  at time  $t$  can be computed as follows:

$$q_i = \frac{1}{t} \sum_{\tau=1}^t (\hat{\theta}_\tau - \bar{s}_{i\tau})^2, \quad (2)$$

where  $\bar{s}_{i\tau}$  is the average of the submitted measurements by user  $i$  at time epochs  $\tau$  prior to  $t$ . Note that we dropped the time index from the quality indicator for notational simplicity.

Quality of service is captured by a generic, positive-valued function  $g(\mathbf{x})$  of participation level vector  $\mathbf{x} = (x_1, \dots, x_N)$ , which includes as parameters the vector of qualities of submitted data by users,  $\mathbf{q} = (q_1, \dots, q_N)$ . Thus, in applications that consider monitoring of a quantity of interest, like air pollution or road traffic condition,  $g(\cdot)$  may denote the *accuracy of estimation*, e.g. in terms of average estimation error. Or, if the application involves decision making, such as detecting whether the measured EMF radiation level through spectrum sensing exceeds a given acceptable threshold,  $g(\cdot)$  may denote the probability of detection or false alarm. We denote by  $\beta$  the level of acceptable quality of service for the subscribers of the application. Hence, the provider needs to operate under the constraint  $g(\mathbf{x}) = \beta$ .

### C. The Mechanism

Upon receiving the request for participation, devices that wish to participate report their perceived cost per unit of

participation level. As mentioned above, this cost is understood as the user offer in terms of minimum requested compensation for participation. The service provider collects user offers in vector  $\mathbf{c} = (c_i : i \in \mathcal{N})$  and it realizes the mechanism.

For given declared cost vector  $\mathbf{c}$ , a mechanism  $M(\mathbf{c})$  consists of computing a *participation level vector*  $\mathbf{x}(\mathbf{c}) = (x_i(\mathbf{c}) : i \in \mathcal{N})$  and a *payment level vector*  $\mathbf{p}(\mathbf{c}) = (p_i(\mathbf{c}) : i \in \mathcal{N})$  for each interested contributor. That is,  $M(\mathbf{c}) = (\mathbf{x}(\mathbf{c}), \mathbf{p}(\mathbf{c}))$ . Note the dependence of participation level  $x_i(\mathbf{c})$  and payment  $p_i(\mathbf{c})$  for each user  $i$  on the entire vector  $\mathbf{c}$  of user declared costs. The provider then announces to each user its participation level and payment. Users respond by submitting data and accordingly get reimbursed.

1) *Bayesian Game*: Once the mechanism is announced to users, a Bayesian game is played. Recall that each user  $i$  knows only its own actual cost  $c_i$  and has only probabilistic information about costs of others. The latter are collectively denoted as  $\mathbf{c}_{-i} = (c_j : j \in \mathcal{N}, j \neq i)$ . Each user  $i$  strategically tries to determine its declared cost in order to maximize its expected utility,

$$\mathbb{E}_{\mathbf{c}_{-i}}[u_i(\mathbf{c})] = \mathbb{E}_{\mathbf{c}_{-i}}[p_i(\mathbf{c}) - c_i x_i(\mathbf{c})], \quad (3)$$

where the expectation above is taken with respect to types of other users. A declared cost vector  $\mathbf{y}^*$  is *Bayesian Nash equilibrium* if for each user  $i \in \mathcal{N}$ ,

$$\mathbb{E}_{\mathbf{y}_{-i}^*}[u_i(y_i^*, \mathbf{y}_{-i}^*)] \geq \mathbb{E}_{\mathbf{y}_{-i}^*}[u_i(y_i, \mathbf{y}_{-i}^*)], \text{ for all } y_i \in \mathcal{C}_i, y_i \neq y_i^*. \quad (4)$$

In other words, in a Bayesian Nash equilibrium, no user has incentive to uni-laterally change its cost declaration strategy, because such a change would not lead to higher utility.

2) *Incentive-compatibility*: A mechanism is called *incentive compatible* (IC), if the strategy where each user reports its true cost is a Bayesian Nash equilibrium, Namely, for each  $i \in \mathcal{N}$ , it is:

$$\mathbb{E}_{\mathbf{c}_{-i}}[p_i(\mathbf{c}) - c_i x_i(\mathbf{c})] \geq \mathbb{E}_{\mathbf{c}_{-i}}[p_i(y_i, \mathbf{c}_{-i}) - c_i x_i(y_i, \mathbf{c}_{-i})], \quad (5)$$

for all  $y_i \in \mathcal{C}_i$  with  $y_i \neq c_i$ , with  $\mathbf{c} = (c_i, \mathbf{c}_{-i})$ . Thus, each user prefers to truthfully report its cost instead of misreporting its cost, given that all other users are truthful.

3) *Individual rationality*: A mechanism is called *individually rational* (IR), if for each  $i \in \mathcal{N}$  and  $c_i \in \mathcal{C}_i$ , it is

$$\mathbb{E}_{\mathbf{c}_{-i}}[u_i(\mathbf{c})] \geq 0, \text{ i.e. } \mathbb{E}_{\mathbf{c}_{-i}}[p_i(\mathbf{c}) - c_i x_i(\mathbf{c})] \geq 0. \quad (6)$$

Individual rationality says that at the Bayesian Nash equilibrium, i.e. the truthful reporting strategy of users, each user has utility at least as much as the one obtained when it does not participate at all. For the latter case, we assume that the participation cost and payment are zero.

### III. PROBLEM STATEMENT AND FORMULATION

The application provider needs to devise a mechanism for participation level and payment allocation to users such that its expected expenditure for reimbursing participants is minimized. The first challenge to confront is user rationality and selfishness. The provider needs to consider the fact that, upon announcing

the mechanism to users, each of them will strategically try to maximize its own utility, i.e. the reimbursement they receive, minus the participation cost.

The second obstacle is that the provider is *unaware* of the actual costs of users. Driven by their strategic behavior, users will try to misreport their true costs in an effort to attract larger reimbursement. Hence, the challenge lies in designing a mechanism, in which the Bayesian Nash equilibrium after user interaction has two desirable properties: (i) it consists precisely of user strategies in which users honestly declare their true cost, because they have no incentive to do otherwise, (ii) users are strongly motivated to participate, i.e. their utility after participation at the Nash equilibrium is greater than zero, which is the utility for no participation.

Third, the provider needs to take into account the different quality of samples provided by users. For that, we need to incorporate in the model the quality indicator discussed above. Finally, the provider needs to conform to delivering a given expected quality of service to its service subscribers.

Define as  $\mathcal{M}(\mathbf{c})$  the space of all mechanisms  $M(\mathbf{c})$  that satisfy the following properties:

- **P1**: The allocation vector  $\mathbf{x}(\mathbf{c})$  satisfies  $g(\mathbf{x}(\mathbf{c})) = \beta$ .
- **P2**:  $M(\mathbf{c})$  is incentive-compatible (IC).
- **P3**:  $M(\mathbf{c})$  is individually rational (IR).

Property P1 refers to the feasibility constraint, namely the mechanism should be such that it provides a given quality of service level  $\beta$ . Properties P2, P3 impose incentive-compatibility and individual rationality respectively.

The problem faced by the service provider is the following:

$$\min_{M(\mathbf{c}) \in \mathcal{M}(\mathbf{c})} \mathbb{E}_{\mathbf{c}} \left\{ \sum_{i \in \mathcal{N}} p_i(\mathbf{c}) \right\}. \quad (7)$$

A. *Conditions for incentive-compatibility and individual rationality*

For each user  $i$ , let  $c_i$  be its true cost and  $y_i$  be the declared cost. Define as  $X_i(y_i)$  the expected allocated participation level to user  $i$ , if  $i$  declares its cost as  $y_i$ , while all other users declare their true costs. That is,

$$X_i(y_i) = \mathbb{E}_{\mathbf{c}_{-i}}[x_i(y_i, \mathbf{c}_{-i})] \quad (8)$$

Also, define as  $P_i(y_i)$  the expected compensation to  $i$ , if it declares cost  $y_i$ , while all other users declare true costs, i.e.,

$$P_i(y_i) = \mathbb{E}_{\mathbf{c}_{-i}}[p_i(y_i, \mathbf{c}_{-i})]. \quad (9)$$

Finally, denote by  $U_i(y_i, c_i)$  the expected utility for user  $i$  if it declares cost  $y_i$  instead of the true  $c_i$ . Clearly, it is

$$U_i(y_i, c_i) = P_i(y_i) - c_i X_i(y_i). \quad (10)$$

Then, the condition for incentive-compatibility, expressed in terms of the quantities above, is

$$U_i(c_i, c_i) \geq U_i(y_i, c_i) \Leftrightarrow P_i(c_i) - c_i X_i(c_i) \geq P_i(y_i) - c_i X_i(y_i) \quad (11)$$

and for individual rationality, it is,

$$U_i(c_i, c_i) \geq 0 \Leftrightarrow P_i(c_i) - c_i X_i(c_i) \geq 0. \quad (12)$$

We now state the following theorem.

*Theorem 1:* A mechanism  $M(\mathbf{c}) = (\mathbf{x}(\mathbf{c}), \mathbf{p}(\mathbf{c}))$  is IC and IR if and only if for all  $i$  the following are true: (a)  $X_i(y_i)$  is non-increasing in  $y_i$ , and (b) The following condition holds:

$$P_i(y_i) = D_i + y_i X_i(y_i) + \int_{y_i}^{\bar{c}_i} X_i(s) ds. \quad (13)$$

where  $D_i = U_i(\bar{c}_i, \bar{c}_i) = P_i(\bar{c}_i) - \bar{c}_i X_i(\bar{c}_i) \geq 0$ , and  $\bar{c}_i$  is the upper limit of the support set of the cost p.d.f.

*Proof:* See Appendix. ■

In the sequel, we can substitute  $y_i$  with  $c_i$  due to incentive-compatibility.

### B. The application service provider optimization problem

We write the objective of the application service provider as

$$\sum_{i \in \mathcal{N}} \mathbb{E}_{\mathbf{c}}[p_i(\mathbf{c})] = \sum_{i \in \mathcal{N}} \mathbb{E}_{\mathbf{c}_i} \{ \mathbb{E}_{\mathbf{c}_{-i}} [p_i(c_i, \mathbf{c}_{-i})] \} = \sum_{i \in \mathcal{N}} \mathbb{E}_{\mathbf{c}_i} \{ P_i(c_i) \}. \quad (14)$$

By substituting  $P_i(c_i)$  from (13) of Theorem 1, we have that for any IC and IR mechanism, we can write each term in the sum in (14) as

$$\begin{aligned} \mathbb{E}_{\mathbf{c}_i} \{ P_i(c_i) \} &= D_i + \int_{\mathcal{C}_i} \left[ c_i X_i(c_i) + \int_{c_i}^{\bar{c}_i} X_i(s) ds \right] f_i(c_i) dc_i \\ &= D_i + \int_{\mathcal{C}_i} c_i X_i(c_i) f_i(c_i) dc_i + \int_{\mathcal{C}_i} \int_{c_i}^{\bar{c}_i} X_i(s) ds f_i(c_i) dc_i. \end{aligned}$$

Consider the integral in the third term above, and use integration by parts, with integration variable  $c_i$  and associated functions  $\int_{c_i}^{\bar{c}_i} X_i(s) ds$  and  $f_i(c_i)$  to write it as

$$\int_{\mathcal{C}_i} X_i(s) ds F_i(c_i) \Big|_{\underline{c}_i}^{\bar{c}_i} - \int_{\mathcal{C}_i} (-X_i(s)) F_i(s) ds = \int_{\mathcal{C}_i} X_i(s) F_i(s) ds, \quad (15)$$

because the first term in the left-hand side above is zero (since  $F_i(\underline{c}_i) = 0$ ), and for the second term we used the fact that  $d(\int_x^a h(x) dx)/dx = -h(x)$  for a real function  $h(\cdot)$ . Thus,

$$\mathbb{E}_{\mathbf{c}_i} \{ P_i(c_i) \} = D_i + \int_{\mathcal{C}_i} c_i X_i(c_i) f_i(c_i) dc_i + \int_{\mathcal{C}_i} X_i(s) F_i(s) ds. \quad (16)$$

Next, we use the definition for  $X_i(\cdot)$  from (8) and we write,

$$\begin{aligned} \mathbb{E}_{\mathbf{c}_i} \{ P_i(c_i) \} &= D_i + \int_{\mathcal{C}_i} c_i \int_{\mathcal{C}_{-i}} x_i(c_i, \mathbf{c}_{-i}) f_{-i}(\mathbf{c}_{-i}) d\mathbf{c}_{-i} f_i(c_i) dc_i \\ &+ \int_{\mathcal{C}_i} \int_{\mathcal{C}_{-i}} x_i(s, \mathbf{c}_{-i}) f_{-i}(\mathbf{c}_{-i}) d\mathbf{c}_{-i} F(s) ds = \\ &= D_i + \int_{\mathcal{C}} c_i x_i(\mathbf{c}) f(\mathbf{c}) d\mathbf{c} \end{aligned} \quad (17)$$

$$\begin{aligned} &+ \int_{\mathcal{C}_i} \int_{\mathcal{C}_{-i}} x_i(s, \mathbf{c}_{-i}) f_{-i}(\mathbf{c}_{-i}) d\mathbf{c}_{-i} \frac{F(s)}{f_i(s)} f_i(s) ds = \\ &= D_i + \int_{\mathcal{C}} x_i(\mathbf{c}) \left( c_i + \frac{F_i(c_i)}{f_i(c_i)} \right) f(\mathbf{c}) d\mathbf{c}, \end{aligned} \quad (18)$$

where  $\mathcal{C} = \times_i \mathcal{C}_i$ ,  $\mathcal{C}_{-i} = \times_{j \neq i} \mathcal{C}_j$ , and  $f(\mathbf{c}) = \prod_i f_i(c_i)$  due to independence of costs. Thus, the total expected compensation cost for the provider is:

$$\sum_{i \in \mathcal{N}} D_i + \sum_{i \in \mathcal{N}} \int_{\mathcal{C}} \left[ x_i(\mathbf{c}) \left( c_i + \frac{F_i(c_i)}{f_i(c_i)} \right) \right] f(\mathbf{c}) d\mathbf{c}. \quad (19)$$

A mechanism  $M(\mathbf{c})$  with  $D_i = 0$  which minimizes

$$\begin{aligned} \int_{\mathcal{C}} \sum_{i \in \mathcal{N}} \left[ x_i(\mathbf{c}) \left( c_i + \frac{F_i(c_i)}{f_i(c_i)} \right) \right] f(\mathbf{c}) d\mathbf{c} &= \\ &= \sum_{i \in \mathcal{N}} \mathbb{E}_{\mathbf{c}} \left[ X_i(\mathbf{c}) \left( c_i + \frac{F_i(c_i)}{f_i(c_i)} \right) \right] \end{aligned} \quad (20)$$

and satisfies properties P1, P2 and P3, solves optimally the problem of the application provider in (7) subject to  $g(\mathbf{x}(\mathbf{c}), \mathbf{q}) = \beta$ , and it is IC and IR.

### C. Optimal mechanism

Consider the following mechanism  $M(\mathbf{c}) = (\mathbf{x}(\mathbf{c}), \mathbf{p}(\mathbf{c}))$  for the provider. For reported cost vector  $\mathbf{c} \in \mathcal{C}$ , let the participation level vector  $\mathbf{x}(\mathbf{c}) = (x_i(\mathbf{c}) : i \in \mathcal{N})$  be the solution to the following optimization problem,

$$\mathbf{x}(\mathbf{c}) = \arg \min_{\mathbf{x}} \sum_{i \in \mathcal{N}} x_i \left( c_i + \frac{F_i(c_i)}{f_i(c_i)} \right) \text{ subject to: } g(\mathbf{x}) = \beta. \quad (21)$$

Also, let the compensation  $p_i(\mathbf{c})$  to each user  $i \in \mathcal{N}$  be

$$p_i(\mathbf{c}) = c_i x_i(\mathbf{c}) + \int_{c_i}^{\bar{c}_i} x_i(s, \mathbf{c}_{-i}) ds. \quad (22)$$

Let  $\delta_i(c_i) = c_i + \frac{F_i(c_i)}{f_i(c_i)} > 0$ . We state the following theorem.

*Theorem 2:* Assume that  $\delta_i(c_i)$  is non-decreasing in  $c_i$ .

(a) If function  $g(\mathbf{x})$  can be written as a monotone function of the sum of terms that are linear in  $x_k$ ,  $k \in \mathcal{N}$ , i.e.

$$g(\mathbf{x}) = h \left( \sum_{k \in \mathcal{N}} \gamma_k x_k \right), \quad (23)$$

with  $\gamma_k \in \mathbb{R}$ , then the mechanism (21), (22) is IC and IR and minimizes the compensation cost of the provider.

(b) If function  $g(\mathbf{x})$  can be written as sum of *concave* strictly increasing functions  $\{g_k(x_k)\}$ ,  $k \in \mathcal{N}$ , i.e.

$$g(\mathbf{x}) = \sum_{k \in \mathcal{N}} g_k(x_k), \quad (24)$$

then the mechanism (21), (22) is IC and IR and minimizes the compensation cost of the provider.

Theorem 2 addresses two potential representative forms of functions  $g(\cdot)$  that capture provisioned quality of experience. The exact form of  $g(\cdot)$  depends on the metric it denotes, and on how data are fused by the provider. Case (a) arises in participatory sensing applications that involve monitoring of a quantity (e.g. air pollution, EMF radiation). Here,  $g(\cdot)$  may denote inaccuracy in reconstructing the underlying field, such as estimation error or probability of detection or false alarm, when data processes from different users are independent. Case (b) emerges in applications which assign a concave valuation

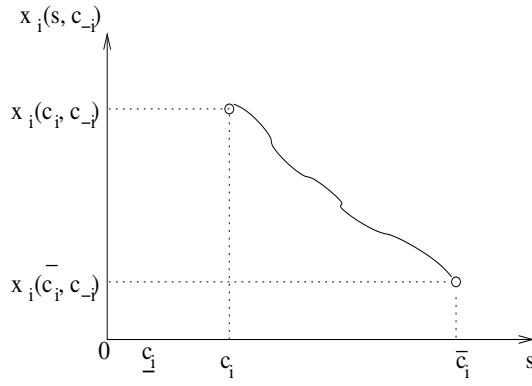


Fig. 2. Participation level  $x_i(\cdot, \mathbf{c}_{-i})$  for user  $i$  as function of user  $i$ 's cost for given  $\mathbf{c}_{-i}$ , and  $\underline{c}_i = 0$ . As discussed above,  $x_i(\cdot, \mathbf{c}_{-i})$  is non-increasing.

function  $g_k(\cdot)$  to data contributed by each user  $k$ . This function captures diminishing returns in the amount of data, and quality of service is measured in terms of the cumulative value. Examples are finding free parking spots in CrowdPark [5] or good biking routes in CycleSense [2].

*Proof:* (a) If  $\mathbf{x}(\mathbf{c})$  minimizes (21) for each  $\mathbf{c}$ , it also minimizes the integral in (20). We show that mechanism (21)-(22) satisfies the conditions of Theorem 1 for being IC and IR, for  $D_i = 0$ .

Since  $P_i(c_i) = \mathbb{E}_{\mathbf{c}_{-i}}[p_i(c_i, \mathbf{c}_{-i})]$ , we use (22) and we get

$$P_i(c_i) = c_i \int_{\mathbf{c}_{-i}} x_i(c_i, \mathbf{c}_{-i}) f_{-i}(\mathbf{c}_{-i}) d\mathbf{c}_{-i} + \int_{\mathbf{c}_{-i}} \int_{c_i}^{\bar{c}_i} x_i(s, \mathbf{c}_{-i}) ds f_{-i}(\mathbf{c}_{-i}) d\mathbf{c}_{-i} = c_i X_i(c_i) + \int_{c_i}^{\bar{c}_i} X_i(s) ds. \quad (25)$$

Thus, condition (13) is satisfied.

To show that  $X_i(\cdot)$  is non-increasing, it suffices to show that  $x_i(c_i, \mathbf{c}_{-i})$  is non-increasing in  $c_i$ , as this will be averaged over  $\mathbf{c}_{-i}$  to imply the same type of monotonicity for  $X_i(\cdot)$ . It can be seen that problem (21) is a linear programming (LP) one. The solution is given by:

$$x_i(c_i, \mathbf{c}_{-i}) = \begin{cases} \frac{h^{-1}(\beta)}{\gamma_i}, & \text{if } \frac{\delta_i(c_i)}{\gamma_i} = \arg \min_{j \in \mathcal{N}} \frac{\delta_j(c_j)}{\gamma_j} \\ 0, & \text{else,} \end{cases} \quad (26)$$

where the inverse  $h^{-1}(\cdot)$  exists due to the monotonicity of  $h(\cdot)$ . It can be seen that if  $\delta_i(c_i)$  is non-decreasing in  $c_i$ , then  $x_i(c_i, \mathbf{c}_{-i})$  is non-increasing in  $c_i$ , and the proof is completed.

(b) The proof goes like that of (a) for showing that (13) is satisfied. To show that  $X_i(\cdot)$  is non-increasing, it suffices again to show that  $x_i(c_i, \mathbf{c}_{-i})$  is non-increasing in  $c_i$ . The Karush Kuhn Tucker condition says that the optimal solution  $\mathbf{x}^*$  of (21) should satisfy:

$$\delta_i(c_i) + \lambda^* g'_i(x_i^*) = 0, \Rightarrow x_i^* = g_i'^{-1} \left( -\frac{\delta_i(c_i)}{\lambda^*} \right) \quad (27)$$

for  $i \in \mathcal{N}$ , where  $g_i'^{-1}(\cdot)$  is the inverse function of  $g_i'(\cdot)$ , and  $\lambda^*$  is the optimal Lagrange multiplier for constraint  $\sum_k g_k(x_k) =$

$\beta$ . Note that  $\lambda^* < 0$ , since  $g'_i(\cdot) > 0$  because  $g_i(\cdot)$  is strictly increasing. Now,  $g'_i(\cdot)$  is non-increasing since  $g_i(\cdot)$  is concave, hence  $g_i'^{-1}(\cdot)$  is non-increasing as well. Also, by assumption,  $\delta_i(c_i)$  is non-decreasing in  $c_i$ . Thus, from (27) we get that  $x_i(\cdot)$  is non-increasing in  $y_i$ , and the proof is completed. ■

The requirement that  $\delta_i(c_i)$  is non-decreasing is met if the pdf  $f_i$  is non-increasing. This is satisfied by a wide range of p.d.fs, including the uniform and exponential ones.

#### D. Interpretation of the mechanism

Consider the mechanism (21), (22). If the provider knew user cost, it would have compensated each user  $i$  with an amount equal to its cost  $c_i$  per unit of participation level; hence the total cost for user  $i$  would be  $c_i x_i(c)$ . However, due to incomplete cost information, the provider gives an extra compensation of  $\int_{c_i}^{\bar{c}_i} x_i(s, \mathbf{c}_{-i}) ds$  in order to motivate them to reveal the actual cost. In order to obtain intuition about the payment mechanism, we define the function

$$T_i(x, \mathbf{c}_{-i}) = \sup\{c \in \mathcal{C}_i : x_i(c, \mathbf{c}_{-i}) \geq x\} \quad (28)$$

This is the maximum cost value that user  $i$  should declare in order to be allocated participation level at least  $x$ , when the other users' cost is  $\mathbf{c}_{-i}$ .

In Fig. 2 we depict an example for the participation level of user  $i$ ,  $x_i(c_i, \mathbf{c}_{-i})$ , as function of its cost  $c_i$ . The compensation in (22) is equal to the area of the rectangle defined by vertical lines at  $s = 0$  and  $s = c_i$  and horizontal lines at 0 and  $x_i(c_i, \mathbf{c}_{-i})$ , plus the area under the curve from  $c_i$  to  $\bar{c}_i$ . Now, note that  $T_i(x, \mathbf{c}_{-i}) = y$  if and only if  $x_i(y, \mathbf{c}_{-i}) = x$ . Thus, the compensation can be written as

$$p_i(\mathbf{c}) = \bar{c}_i x_i(\bar{c}_i, \mathbf{c}_{-i}) + \int_{x_i(\bar{c}_i, \mathbf{c}_{-i})}^{x_i(c_i, \mathbf{c}_{-i})} T_i(x, \mathbf{c}_{-i}) dx. \quad (29)$$

That is, each user  $i$  is first compensated with  $\bar{c}_i x_i(\bar{c}_i, \mathbf{c}_{-i})$ ; on top of that, for each additional unit of increase of participation level from  $x$  to  $x+dx$ , the user is compensated with an amount  $T_i(x, \mathbf{c}_{-i})$ , equal to the maximum cost he would declare to get participation level  $x$ , when other users' costs are  $\mathbf{c}_{-i}$ .

In order to compute the participation level allocation and payments, the provider first solves (21), which may be linear or convex. Next, it needs to compute for each user  $i$  the tentative allocations  $x_i(s, \mathbf{c}_{-i})$  for  $s \in [\underline{c}_i, \bar{c}_i]$  so as to compute the payments in (22). In practice, it may quantize the interval above and solve a set of optimization problems for each user  $i$ . Likewise, it may approximate the integral as a summation and employ numerical methods.

#### IV. CASE STUDY: PARAMETER ESTIMATION

As a case study, we consider a participatory sensing application that involves accurate estimation of a slowly time-varying, unknown, spatially homogeneous process in a certain area (e.g. air pollution or EMF radiation at that area). Time is divided in epochs. In order to estimate the value  $\theta$  of the process in each epoch, the provider sends a request for data contribution and receives offers by  $N$  sensors. Fix attention to the auction at one epoch.

Device  $i$  takes several measurements  $\{z_i(\tau)\}$  at times  $\tau$  in the epoch, given by  $z_i(\tau) = \theta + n_i(\tau)$ . The noise process  $n_i(\tau)$  captures uncertainty of  $i$ 's measurement due to different perception of the phenomenon process, or due to residual measurement errors. For each  $i$ ,  $n_i$  can be Gaussian, zero mean and stationary. The variance of  $n_i$ ,  $\sigma_i^2 = \mathbb{E}[n_i^2]$ , captures measurement inaccuracy. Noise processes of any two sensors  $i$  and  $j$  are spatially and temporally uncorrelated. Sensors submit measurements to the provider. At the end of the epoch, the provider computes an estimate  $\hat{\theta}$  of the unknown parameter  $\theta$  in the Maximum Likelihood sense and transmits the result to service subscribers. Suppose that sensor  $i$  takes  $x_i$  measurements. The criterion for quality of service is mean squared estimation error (MSEE),  $\mathbb{E}[(\theta - \hat{\theta})^2]$ . In [22], we have shown that the MSEE is  $g(\mathbf{x}) = \left(\sum_{i=1}^N \frac{x_i}{\sigma_i^2}\right)^{-1}$ .

The quality  $q_i$  of measurement data of sensor  $i$  can be captured by variance,  $\sigma_i^2$ . However, the provider does not know  $\sigma_i^2$  and uses as approximation the average squared deviation of data of sensor  $i$  from computed estimates  $\{\hat{\theta}_\tau\}$  at previous epochs  $\tau$ . Namely,  $\sigma_i^2$  is approximated by  $q_i$  in (2). The participation level of sensor  $i$  is the number of its measurements,  $x_i$ . The provider has the constraint  $g(\mathbf{x}) = \varepsilon$ , where  $\varepsilon$  is a specified acceptable estimation error, and thus,

$$\sum_{i=1}^N \frac{x_i}{q_i} = \frac{1}{\varepsilon}. \quad (30)$$

The cost  $c_i$  (e.g. energy cost) per measurement for sensor  $i$  is unknown to the provider and is assumed to be uniformly distributed in  $[\alpha_i, b_i]$ . The coefficient of  $x_i$  is  $\delta_i(c_i) = c_i + \frac{F_i(c_i)}{f_i(c_i)} = 2c_i - \alpha_i$ . The mechanism that minimizes the compensation cost for the provider subject to the quality of service constraint (30) and being IC and IR, comes from the solution to (21). In our case, the problem is an LP one. Let

$$i^* = \arg \min_{i=1,\dots,N} (2c_i - \alpha_i)q_i. \quad (31)$$

Then, the optimal participation levels are:

$$x_{i^*} = \frac{q_{i^*}}{\varepsilon}, \text{ and } x_j = 0 \text{ for all } j \neq i^*. \quad (32)$$

Define function  $z(\mathbf{c}_{-i}) = \sup\{c : q_i \delta_i(c) \leq \min_{k \neq i} q_k \delta_k(c_k)\}$ . This is the maximum declared cost of  $i$  that can make him win the auction against declared costs of others, and it is

$$z(\mathbf{c}_{-i}) = \frac{1}{2} \left[ \alpha_i + \frac{1}{q_i} \min_{k \neq i} [q_k (2c_k - \alpha_k)] \right]. \quad (33)$$

Rule (32) can be written as:

$$x_i(\mathbf{c}_{-i}) = \begin{cases} q_i/\varepsilon, & \text{if } c_i \leq z(\mathbf{c}_{-i}), \\ 0, & \text{else.} \end{cases} \quad (34)$$

From (22), the payment for the single selected user will be:

$$p_i = \frac{q_i}{\varepsilon} z_i(\mathbf{c}_{-i}). \quad (35)$$

The participation level allocation mechanism takes into account declared costs, lower bounds of cost intervals and data qualities. Among sensors with the same data quality, the

mechanism chooses sensor  $i$  with the smallest  $(2c_i - \alpha_i)$ . Interestingly enough, the mechanism does not simply favour the sensor with the smallest cost  $c_i$  but also the one for which the cost is closer to the lower bound. Among sensors with the same  $(2c_i - \alpha_i)$ , the sensor with the smallest  $q_i$  (i.e the best data quality) is selected. The participation level and payment are proportional to the data quality of the selected sensor.

The VCG mechanism [23, Chap.6], [24, Sec.II.C], which leads to socially optimal allocation while being IC and IR, allocates participation level and compensation so as to minimize total sensor cost,  $\sum_{i=1}^N c_i x_i$ , subject to (30).

Let  $\tilde{i} = \arg \min_i c_i q_i$ , then the VCG participation level selection is:  $x_{\tilde{i}} = q_{\tilde{i}}/\varepsilon$ , and  $x_j = 0$ , for  $j \neq \tilde{i}$ . The VCG compensation for the winner  $\tilde{i}$  is the difference between the total cost of sensors  $j \neq \tilde{i}$  when winner  $\tilde{i}$  does not participate in the auction, and the total cost of sensors  $j \neq \tilde{i}$  when  $\tilde{i}$  participates in the auction. In the latter case, this cost is zero. Now, if the winner  $\tilde{i}$  does not participate in the auction, there will be another sensor,  $i'$  with the second smallest  $c_i q_i$  who will be winner and will be allocated all participation level; then the total cost for sensors  $j \neq \tilde{i}$  is  $c_{i'} q_{i'}/\varepsilon$ . Hence, the compensation to winner  $\tilde{i}$  is  $c_{i'} q_{i'}/\varepsilon$ .

## V. CONCLUSION

We addressed the design of incentive mechanisms for participatory sensing applications that are optimal in the sense of minimizing compensation cost to participants by the application provider, subject to delivering a certain quality of service to subscribers. Inherent in our design are incentives for user participation and truthful cost declaration. The model we considered includes the participation cost and compensation for each user and serves as a first step towards understanding the structure of the solution. Future research could enhance the model along various directions, such as considering the similarity of device measurements (e.g. due to neighboring locations) when deciding about participation levels and payments. This could further improve the cost of the provider, while eliminating potential data redundancies.

## VI. ACKNOWLEDGMENT

The author acknowledges the support of ERC08-RECITAL project, co-financed by Greece and the European Union (European Social Fund) through the Operational Program "Education and Lifelong Learning" - NCRF 2007-2013.

## REFERENCES

- [1] R.K. Ganti, F. Ye, H. Lei, "Mobile Crowdsensing: Current state and future challenges", *IEEE Comm. Mag.*, vol.49, no.11, pp.32-39, Nov. 2011.
- [2] J. Goldsman, K. Shilton, J. Burke, D. Estrin, M. Hansen, N. Ramanathan, S. Reddy, V. Samanta, M. Srivastava, "Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world", *White Paper*, Center for Embedded Networked Sensing, UCLA, 2008.
- [3] OpenSense Project: <http://www.nano-tera.ch/projects/401.php>, 2010.
- [4] R.K. Ganti, N. Pham, H. Ahmadi, S. Nangia, and T.F. Abdelzaher, "GreenGPS: A Participatory sensing fuel-efficient maps application", in *Proc. ACM MobiSys*, 2010.
- [5] T. Yan, B. Hoh, D. Ganesan, K. Tracton, T. Iwuchukwu, J.-S. Lee, "CrowdPark: A Crowdsourcing-based Parking Reservation System for Mobile Phones", *University of Massachusetts at Amherst Tech. Report*, <http://lass.cs.umass.edu/yan/pubs/yan11CrowdPark.pdf>



- [6] S. Mathur, T. Jin, N. Kasturirangan, J. Chandrashekarhan, W. Xue, M. Gruteser, and W. Trappe, "Parknet: Drive-by sensing of road-side parking statistics.", in *Proc. of ACM MobiSys*, 2010, pp. 1231-1236.
- [7] L. Deng and L.P.Cox, "LiveCompare: Grocery Bargain Hunting through Participatory Sensing", in *Proc. ACM HotMobile*, 2009.
- [8] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin and M. Hansen, "Image browsing, processing, and clustering for participatory sensing: Lessons from a dietsense prototype", in *Proc. 4th Workshop on Embedded Net. Sensors (EmNets)*, 2007.
- [9] A. Krause, E. Horvitz, A. Kansal and F. Zhao, "Toward community sensing", in *Proc. ACM/IEEE Int. Conf. on Inform. Proc. in Sensor Networks (IPSN)*, 2008.
- [10] S. Reddy, D. Estrin, M. Hansen, and M. Srivastava, "Examining micro-payments for participatory sensing data collections", in *Proc. ACM Int. Conf. on Ubiquitous computing (Ubicomp)*, 2010.
- [11] C. H. Liu, P. Hui, C. Bisdikian, J. W. W. Branch, and B. Yang, "Efficient network management for context-Aware participatory sensing", in *Proc. IEEE SECON*, 2011.
- [12] T. Luo and C.-K. Tham, "Fairness and social welfare in incentivizing participatory sensing", in *Proc. IEEE SECON*, 2012.
- [13] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications", in *Proc. IEEE INFOCOM*, 2012.
- [14] R. Gibbons, *Game Theory for Applied Economists*, Princeton University Press, 1992.
- [15] V. Krishna, *Auction Theory*, Academic Press, 2002.
- [16] W. Vickrey, "Counterspeculation, Auctions and Competitive Sealed Tenders", *Journal of Finance*, vol. 16, pp. 8-37, 1961.
- [17] T. Groves, "Incentives in Teams", *Econometrica*, vol. 41, pp. 617-631, 1973.
- [18] R. Myerson, "Optimal auction design", *Math. of Oper. Research*, vol.6, pp.58-73, 1981.
- [19] D. Yang, G. Xue, X. Fang and J. Tang, "Crowdsourcing to Smartphones: Incentive mechanism design for mobile phone sensing", in *Proc. ACM MobiCom*, 2012.
- [20] J.-S. Lee and B. Hoh, "Dynamic pricing incentive for participatory sensing", *Elsevier Pervasive and Mobile Computing*, vol.6, no.6, pp. 693-708, Dec. 2010.
- [21] X. Zhuo, W. Gao, G. Cao and Y. Dai, "Win-Coupon: An incentive framework for 3G traffic offloading", in *Proc. IEEE Int. Conf. on Net. Protocols (ICNP)*, 2011.
- [22] I. Koutsopoulos and M. Halkidi, "Lifetime maximization in wireless sensor networks with an estimation mission", in *Proc. IEEE GLOBECOM*, 2009.
- [23] S. Shakkottai and R. Srikant, *Network Optimization and Control*, Foundations and Trends in Networking, NoW Publishers, 2007.
- [24] S. Yang and B. Hajek, "VCG-Kelly mechanisms for allocation of divisible goods: Adapting VCG mechanisms to one-dimensional signals", *IEEE J. Selected Areas of Communications*, vol. 25, no.6, pp. 1237-1243, 2007.

## APPENDIX

## PROOF OF THEOREM 1

First, assume that for a mechanism  $M(\mathbf{c})$ ,  $X_i(\cdot)$  is non-increasing and condition (13) holds. We will show that  $M(\mathbf{c})$  is IC and IR for any user  $i$ . For  $c_i \in \mathcal{C}_i$  we have

$$U_i(c_i, c_i) = P_i(c_i) - c_i X_i(c_i) = D_i + \int_{c_i}^{\bar{c}_i} X_i(s) ds \geq 0, \quad (36)$$

where we used (13), and the fact that  $D_i \geq 0$  and  $X_i(\cdot) \geq 0$ . Thus, the mechanism is IR. Furthermore by using again (13),

$$U_i(c_i, c_i) - U_i(y_i, c_i) = \int_{c_i}^{\bar{c}_i} X_i(s) ds - \int_{y_i}^{\bar{c}_i} X_i(s) ds + (c_i - y_i) X_i(y_i). \quad (37)$$

We distinguish two cases: (i) if  $y_i < c_i$ , then for the right-hand side of (37) we have

$$\begin{aligned} & - \int_{y_i}^{c_i} X_i(s) ds + (c_i - y_i) X_i(y_i) \geq \\ & - X_i(y_i) \int_{y_i}^{c_i} ds + (c_i - y_i) X_i(y_i) = 0, \end{aligned}$$

where the inequality is due to the fact that  $X_i(s) \leq X_i(y_i) \forall s \in [y_i, c_i]$ , since  $X_i(\cdot)$  is non-increasing.

(ii) If  $c_i < y_i$ , the right-hand side of (37) becomes

$$\begin{aligned} & \int_{c_i}^{y_i} X_i(s) ds + (c_i - y_i) X_i(y_i) \geq \\ & X_i(y_i) \int_{c_i}^{y_i} ds + (c_i - y_i) X_i(y_i) = 0, \end{aligned}$$

where the inequality is due to the fact that  $X_i(s) \geq X_i(y_i) \forall s \in [c_i, y_i]$ , since  $X_i(\cdot)$  is non-increasing. Thus, we have proved that  $U_i(c_i, c_i) \geq U_i(y_i, c_i)$  for all  $c_i, y_i \in \mathcal{C}_i$ , and thus the mechanism is IR for each user  $i$ .

For the inverse, consider an IC and IR mechanism. We will prove conditions (a)-(b). Incentive-compatibility implies that for any  $w, z \in \mathcal{C}_i$  with  $w < z$ , it is  $P_i(w) - w X_i(w) \geq P_i(z) - w X_i(z)$  and  $P_i(z) - z X_i(z) \geq P_i(w) - z X_i(w)$ . By adding these inequalities and rearranging terms, we get  $(z - w) X_i(z) \leq (z - w) X_i(w)$ . Since  $z - w > 0$ , it must be  $X_i(z) \leq X_i(w)$ , and thus  $X_i(\cdot)$  is non-increasing, thus (a) is proved.

For condition (b), we start by defining  $G_i(c_i) = U_i(c_i, c_i)$ . Due to incentive-compatibility, it is:

$$G_i(c_i) = \max_{y_i \in \mathcal{C}_i} U_i(y_i, c_i) = \max_{y_i \in \mathcal{C}_i} \{P_i(y_i) - c_i X_i(y_i)\}. \quad (38)$$

This implies that  $G_i(c_i)$  is the maximum of a family of affine functions of  $c_i$ , and thus it is convex and differentiable everywhere, except at countably many points. We compute

$$\begin{aligned} \lim_{\epsilon \rightarrow 0} \frac{G_i(c_i + \epsilon) - G_i(c_i)}{\epsilon} & \geq \lim_{\epsilon \rightarrow 0} \frac{U_i(c_i, c_i + \epsilon) - G_i(c_i)}{\epsilon} = \\ \lim_{\epsilon \rightarrow 0} \frac{P_i(c_i) - (c_i + \epsilon) X_i(c_i) - P_i(c_i) + c_i X_i(c_i)}{\epsilon} & = -X_i(c_i) \end{aligned} \quad (39)$$

where the inequality above is due to (38). Similarly,

$$\lim_{\epsilon \rightarrow 0} \frac{G_i(c_i) - G_i(c_i - \epsilon)}{\epsilon} \leq -X_i(c_i) \quad (40)$$

Inequalities (39) and (40) mean that  $G'_i(c_i) = -X_i(c_i)$ . Therefore, for any  $y_i \in \mathcal{C}_i$ ,

$$G_i(\bar{c}_i) - G_i(y_i) = - \int_{y_i}^{\bar{c}_i} X_i(s) ds \quad (41)$$

and by substituting  $G_i(\cdot)$ , we have

$$P_i(\bar{c}_i) - \bar{c}_i X_i(\bar{c}_i) - P_i(y_i) + y_i X_i(y_i) = - \int_{y_i}^{\bar{c}_i} X_i(s) ds, \quad (42)$$

and finally

$$P_i(y_i) = D_i + y_i X_i(y_i) + \int_{y_i}^{\bar{c}_i} X_i(s) ds. \quad (43)$$

Thus, we get (13) of Theorem 1, and we proved condition (b). Note that individual rationality at  $\bar{c}_i$  implies  $P_i(\bar{c}_i) - \bar{c}_i X_i(\bar{c}_i) \geq 0$ , which means  $D_i \geq 0$ .