

# Incentive Allocation to Sequential Decision-Making Sensors in Mobile Crowdsensing

Iordanis Koutsopoulos  
Department of Informatics  
Athens University of Economics and Business  
Athens, Greece  
Email: jordan@aueb.gr

**Abstract**—In this work in progress, we consider incentive allocation to a set of measurement sensors in the context of mobile crowdsensing. The novelty stems from considering a new model perspective for each sensor, that of a rational sequential decision-maker. At each time slot, each sensor observes the time-varying cost it undergoes for submitting measurements and the advertised reward for submitting measurements to the platform. Its decision policy at each time slot is whether to become active and submit measurements or stay inactive. The sensor decision problem is shown to be described as an optimal stopping one, and the sensor policy that maximizes its expected net benefit over a time horizon is shown to be of threshold nature at each time slot, where the threshold is non-increasing with the elapsed time. With the derived optimal policies for sensors, we next seek to determine the optimal price per time slot paid by the platform to each sensor so as to maximize the expected total quality of collected measurements, subject to a budget constraint. Finally, we introduce the problem of centralized sensor activation in a dynamically varying system so as to maximize the long-term average utility stemming from the quality of collected data. The characterization of distributed sensor equilibrium policies and the assessment of their impact on the global performance metric compared to the optimal centralized policy, are outlined as important directions that warrant further investigation.

**Index Terms**—Mobile crowdsensing; sequential decision-making; optimal stopping problem; incentive allocation.

## I. INTRODUCTION

Mobile crowdsensing is a flourishing paradigm in pervasive and mobile computing, which features data collection from multiple embedded sensors on mobile devices and wearables, and subsequent data processing and analytics. The interface with users is oftentimes realized through a mobile app, and ultimately the outcome of data processing is delivered to mobile users as a service through the app. Applications span areas such as environmental monitoring, urban sensing and reporting, transportation, healthcare, lifestyle improvement, or citizen journalism [1].

The viability of provisioning the service empowered by mobile crowdsensing depends crucially on the participation of mobile devices and their users. Devices submit data through their sensors subject to various resource limitations of mobile devices (e.g. mobile device battery cost for sensing, processing and transmission, or device processing power cost), monetary

cost paid to the mobile operator for transmitted data or simply user attention and time due to the effort to submit data. It is therefore important to have incentive mechanisms to motivate users to contribute their measurements in the pool of collected data. Incentives are often provided as direct monetary reimbursements or in the form of various offers or discounts, and they constitute a prime means of implicitly controlling the set of distributed sensors towards achieving the quality of service goals.

In order to design proper incentives, it is important to accurately model involved mobile devices and users and their interaction with the online platform. Namely, it is crucial to take into account the fact that embedded sensors on mobile devices act rationally towards optimizing their own objective through the software that controls their decisions.

### A. Related work

In mobile crowdsensing applications, data measurements are submitted either automatically without user intervention with sensors collecting data in the background, or through direct participation of users. In the former case, when the selection and compensation of sensors by the platform takes place through an auction, sensors act strategically in reporting their true cost of providing their data in an effort to get larger reimbursement and maximize their net benefit. In that case, mechanism design through optimal auction theory is employed to select sensors and determine payments such that true cost reporting is a Bayesian Nash equilibrium [2].

A Stackelberg game formulation is considered in [3], in which the platform chooses an appropriate total reward for a task and a threshold number of devices to share it, and devices choose whether to participate or not in the task. Another Stackelberg model formulation is presented in [4], where the platform announces a total reward, and sensors determine their sensing plan in terms of sensing times by taking into account the sensing cost and the fact that the total reward is allocated to sensors in proportion to their sensing times. The platform determines the total reward through the maximization of its net utility i.e. the benefit minus the reward, where the benefit is a logarithmic function of sensor sensing times that emerge from Nash equilibrium.

The provision of incentives has been investigated across a longer-time scale as well. In [5], the authors use the framework

This work was supported by the European Commission H2020 Research Program under Grant Number 688768 - netCommons (Network Infrastructure as Commons).

of Lyapunov optimization to design an online algorithm for sensor selection at each time slot for maximizing social welfare, which is defined as the total sensing value minus the sensing cost. The long-term participation of users is achieved by ensuring that the probability of selecting each of them is no smaller than a threshold. The work [6] gives a comprehensive survey on incentive design for mobile crowdsensing.

On the other hand, when users are actively involved in data submission, the assumption of user/device rationality is challenged. In that case, users' choice can be better captured through machine-learning models. In [7], we modeled each user through a logistic-regression model that determines the probability that the user will respond to a task request or not based on two attributes, the distance traveled or effort exerted to execute the task, and the incentive payment. Finally, a different perspective on incentives for mobile crowdsourcing is presented in [8], where a gamification approach attempts to boost user engagement via user ranking and status level schemes based on reward points.

Optimal stopping game formulations have appeared previously in several contexts. In [9], the author considers a Markov stopping game between two competing players in the context of resource sharing. Players monitor the same sequence that characterizes the status of the opportunistic resource and make stopping decisions independently in order to obtain the resource. The player that stops first reaps the current reward and leaves the system, while the other player solves a classical optimal stopping problem for the remaining time horizon. The framework is extended to more than two players in [10]. Another application of optimal stopping in resource sharing is studied in [11]. Here, each agent decides whether to stay in one location or move to another one after monitoring the resource level and number of agents at each location. The authors study the equilibrium behavior of agents in the limit of large number of agents and locations and show that it has a threshold nature, namely that an agent decides to switch to another location based only on the current locations resource level and the number of agents.

### B. Our contribution

Common to most of the existing literature is the assumption that the device utility and decisions are abstracted and guided by static analytical utility functions, and the goal is to design a mechanism that optimizes a certain objective such as minimizing the payment expenses, maximizing the quality of collected information or maximizing the profit of the platform.

In this work, we focus on the class of mobile crowdsensing scenarios that require continuous supply of sensor measurement data, e.g. for monitoring or other purposes. In these scenarios, modeling the sensors and their decisions through static utility functions fails to capture sequential sensor decisions about participation in crowdsensing tasks since these need to be taken on the fly as the data collection process evolves and the dynamics of the system change. These decisions are taken by the software engine that controls sensor operation after considering the time-varying measurement cost or other ambient

conditions, and they lead to sequential decision-making rules for sensors. Although some works discuss sequential auctions or dynamic games (see e.g. [6] and references therein), they do not capture the true impact of time-varying conditions and do not model a sensor sequential decision policy as a means of reacting to these conditions.

We attempt a radical departure from the aforementioned approaches by explicitly modeling the sensor as a rational sequential decision maker which decides on its activation or deactivation at each slot based on sequential observations, with the goal to optimize its expected net benefit. Subsequently, we ask the question: how can the platform influence the sequential decision-making process of sensors through incentive design, towards the global goal it wants to achieve, namely maximizing the average quality of collected data?

The contributions of this work in progress are as follows.

- We formulate the sequential decision problem of each sensor about being active and participating in data submission as an optimal stopping problem. At each time slot, each sensor observes the time-varying cost it undergoes for submitting measurements, and the advertised reward. Its decision policy at each time slot is whether to become active and submit measurements in that slot or to stay inactive and idle. The sensor decision problem is formulated as an optimal stopping one, and the sensor policy that maximizes its net benefit over a time horizon is shown to be of threshold nature at each time slot, where the threshold is non-increasing with elapsed time.
- With the derived optimal policies for sensors, we next formulate the problem of determining the optimal price per time slot paid by the platform to each sensor so as to maximize the expected total quality of collected measurements, subject to a budget constraint.
- Finally, we introduce the problem of centralized sensor activation in a dynamically varying system so as to maximize the long-term average utility stemming from the quality of collected data. We outline various important directions for further investigation such as the characterization of sensor equilibrium policies and the assessment of their impact on the global performance metric compared to the optimal centralized policy.

The rest of the paper is organized as follows. In section II, we present the model and formulate the single-sensor problem as an optimal stopping one. In section III, we set the stage for the problem of price allocation, and for the comparison of individual sensor policies and the centralized policy in a dynamically varying system. We conclude in section IV.

## II. MODEL

A service provider wishes to collect measurement data from a set  $\mathcal{N}$  of  $n$  devices (sensors) in an area of interest. It coordinates a mobile crowdsensing platform over a time horizon of  $T$  equal-sized time slots of unit duration. The durations of the horizon and the time slot are dictated by the application. For example, the horizon duration may be in the

range of several hours or days, while the time slot duration may be from seconds to few minutes or more.

At the beginning of each time slot  $k$ ,  $k = 0, 1, \dots, T-1$ , each sensor  $i$  observes or assesses the value of its cost  $c_{i,k}$  for submitting measurements with a fixed rate  $r$  during this slot, thus for submitting  $r$  measurements per slot. This cost may be associated to energy consumption, bandwidth consumption or computational processing cost, or their cumulative effect. For example, it might capture the different energy cost to submit to the nearest WiFi access point or macro-cellular base station. It may also represent the impact of the measurement collection process on the smartphone processor depending on other concurrent running processes. Or it could capture the level of discomfort to the user, depending on her current activity. The cost is time-varying and assumed to be i.i.d across time slots. Let  $C_i$  be the random variable that models the random cost of sensor  $i$ , with known distribution and expected value  $\mathbb{E}[C_i]$ .

The service provider advertises a payment  $p_i$  per time slot to each sensor  $i$ . This is fixed for the entire horizon, it is determined a priori, and it is announced to each device at time  $k = 0$ .

#### A. Single-sensor decision problem

If a sensor  $i$  is inactive at  $k = 0$ , it may switch to being active at some time slot  $\tau$ , with  $0 < \tau \leq T-1$  during the time horizon. On the other hand, sensors that are active at  $k = 0$  may switch to being inactive at some time  $\tau'$  with  $0 < \tau' \leq T-1$  during the horizon.

In order to better demonstrate the problem faced by each sensor, we start with the assumption that each sensor performs exactly one switch in the horizon of  $T$  slots, in the worst case in the last time slot of the horizon. This assumption can be relaxed without loss of generality so that sensors may perform multiple switches between being inactive and being active during the horizon. Indeed, when a sensor switches from inactive to active at time  $\tau$ , it faces a new problem with time horizon of  $T - \tau$  during which it should decide whether it should become inactive again. When it switches to being inactive at some time  $\tau' \in \{0, \dots, T - \tau\}$ , it faces a new problem with horizon  $T - \tau - \tau'$  during which it should decide when to become active, and so on. If the sensor is active in a slot, it submits  $r$  measurements.

The state at a slot  $k$  is the sensor measurement/transmission cost  $c_{i,k}$ . In the sequel, we drop index  $i$  from  $c_{i,k}$ , random variable  $C_i$  and  $p_i$ , and we denote the cost at time slot  $k$  as  $c_k$ , the random variable as  $C$ , and the payment as  $p$ .

1) *Inactive sensors at  $k = 0$* : The problem faced by a sensor falls under the category of optimal stopping problems [12, Sec.4.4]. Fix attention to a sensor that is inactive at time  $k = 0$ . At each time slot  $k$  the sensor observes or calculates cost  $c_k$  and needs to decide whether to continue to be inactive or to activate itself and submit measurements. Note that there is a (fixed) energy cost for being active, but without loss of generality we absorb it to cost  $c_k$ . If the node decides to be inactive, it undergoes 0 cost. On the other hand, if it decides

to activate itself in slot  $k$ , it contributes  $r$  measurements in that slot. Thus, it experiences cost  $c_k$  and receives payment  $p$ , and it also expects to experience an expected net payoff-to-go until the end of the horizon. The latter is because, once a node is activated, it will continue to be active and contribute measurements for the rest of the horizon. Define  $J_k(c_k)$  to be the *maximum expected net-payoff-to-go* for time slot  $k$ .

Since the sensor may switch to being active at most once during the horizon, in the extreme case it will switch at the beginning of slot  $T$  with a net benefit  $J_T(c_T) = p - c_T$ . At the beginning of slot  $T-1$ , the sensor faces the dilemma to continue being inactive and experience net payoff zero or become active. In the latter case, it will experience a net payoff at that slot equal to  $p - c_{T-1}$  and also an expected net payoff-to-go,  $\mathbb{E}[J_T(c_T)]$  for slot  $T$ . where the expectation is with respect to randomness of the cost.

The optimal expected net payoff-to-go is written as

$$\begin{aligned} J_{T-1}(c_{T-1}) &= \max\{0, p - c_{T-1} + \mathbb{E}[J_T(c_T)]\} \\ &= \max\{0, 2p - c_{T-1} - \mathbb{E}[C]\}, \end{aligned} \quad (1)$$

and therefore the optimal decision for the sensor is:

$$\begin{aligned} &\text{activate, if } c_{T-1} \leq 2p - \mathbb{E}[C] \\ &\text{stay inactive, if } c_{T-1} > 2p - \mathbb{E}[C]. \end{aligned}$$

Similarly, for any time slot  $k = 0, \dots, T-1$ , it is

$$J_k(c_k) = \max\{0, p - c_k + \mathbb{E}[J_{k+1}(c_{k+1})]\}. \quad (2)$$

and the optimal decision is

$$\begin{aligned} &\text{activate if } c_k \leq p + (T - k + 1)(p - \mathbb{E}[C]) \\ &\text{stay inactive, if } c_k > p + (T - k + 1)(p - \mathbb{E}[C]) \end{aligned} \quad (3)$$

after making the meaningful assumption that  $p > \mathbb{E}[C]$ . Note that the switch decision at each time slot  $k$  is of a threshold form, i.e. the sensor switches when  $c_k \leq \alpha_k$  where threshold  $\alpha_k$  is non-increasing with  $k$ . This implies that as time elapses, the switch is less likely to happen.

2) *Active sensors at  $k = 0$* : Now, consider a sensor that is active at  $k = 0$ . At each time  $k$ , it decides whether it will continue to be active and contribute measurements or whether it will switch off. This will happen at the latest in the last slot, thus  $J_T(c_T) = 0$ . At slot  $T-1$  we have

$$J_{T-1}(c_{T-1}) = \max\{0, p - c_{T-1}\} \quad (4)$$

and the optimal decision is

$$\begin{aligned} &\text{become inactive, if } c_{T-1} \leq p \\ &\text{stay active, if } c_{T-1} > p. \end{aligned}$$

Also

$$J_{T-2} = \max\{0, p - c_{T-2} + \mathbb{E}[\max\{0, p - c_{T-1}\}]\}. \quad (5)$$

and the optimal decision is

$$\begin{aligned} &\text{become inactive, if } c_{T-2} \leq 2p - \mathbb{E}[C] \\ &\text{stay active, if } c_{T-2} > 2p - \mathbb{E}[C]. \end{aligned}$$

and in general, the optimal decision is

$$\begin{aligned} &\text{become inactive, if } c_k \leq p + (T - k)(p - \mathbb{E}[C]) \\ &\text{stay active, if } c_k > p + (T - k)(p - \mathbb{E}[C]). \end{aligned} \quad (6)$$

### III. FUTURE DIRECTIONS

We outline two important directions for future research. The first concerns optimal incentive allocation in a static scenario where the probabilities that sensors are active are computed through the stopping policies of section II. The second setting is about comparing the centralized and autonomous sensor activation policies in a dynamic setting.

#### A. Optimal incentive payment allocation

Given the aforementioned sequential decision-making of rational sensors, the platform aims to find the incentive payment policy  $\mathbf{p} = (p_i : i = 1, \dots, n)$ , where  $p_i$  is the payment to sensor  $i$ , so as to guide sensors towards activation policies that are desirable for the system as a whole. Each sensor has a quality index  $g_i$  about the average quality of its measurements.

The platform operator does not have access to the sequence of costs  $\{c_{i,k}\}_{k=0,1,\dots}$  of each sensor  $i$ , but it may know the empirical distribution of costs based on historical data. With enough historical data, the empirical distribution may approximate well  $\mathbb{E}[C_i]$  for each sensor  $i$ . The platform operator can then compute the probability that a node is on at a time slot.

Let  $\alpha_{i,k}$  denote the threshold that dictates the switching for a sensor  $i$  at time slot  $k$ . If a sensor  $i$  was inactive at  $k = 0$ , the probability that it is active in slot  $k$  is

$$\Pr[\text{sensor } i \text{ is active in time slot } k] = \sum_{\ell=1}^k q_{i,\ell} \prod_{m=1}^{\ell-1} (1 - q_{i,m}) \quad (7)$$

where

$$q_{i,\ell} = \Pr[C_i \leq \alpha_{i,\ell}]. \quad (8)$$

A similar expression holds for sensors that were active at slot  $k = 0$ . The objective of the platform is to maximize the total quality of collected measurements subject to budget constraints. If the total quality is additive in measurements, the objective can be stated as follows,

$$\max_{\mathbf{p}} \sum_{k=1}^T \sum_{i=1}^n g_i \Pr[\text{sensor } i \text{ is active in time slot } k] \quad (9)$$

subject to

$$\sum_{i=1}^n p_i \leq B, \quad (10)$$

where  $B$  is the total budget available for incentive payments. Another constraint might be that the gathered data should have sufficient quality, e.g. at least  $G$ , at each time slot. Intuitively, the platform tries to calibrate the decisions of sensors by providing them with appropriate payments so that it achieves a balancing of active sensors across time slots, so that a desired quality level,  $G$  is achieved at each time slot.

#### B. Impact of multiple sequential decision makers on a global performance objective in a dynamic setting

We now consider a dynamic scenario where sensors arrive according to a (Poisson) arrival process with average rate  $\lambda$  sensors per unit of time. A sensor is available for a random duration which is exponentially distributed with parameter  $s$ , i.e.  $\Pr(s \leq x) = 1 - e^{-sx}$ ,  $x \geq 0$ . Equivalently, the average duration of availability for a sensor is  $1/s$  time units, and  $s$  is the average rate with which sensors leave the system. Although these assumptions are motivated for better mathematical tractability as they facilitate the derivation of the optimal policy, they are close to reality since they model the burst of arriving and departing sensors from the system.

At time slot  $t$ , let  $W(t)$  be the aggregate quality of data provided by all active sensors in that slot. For simplicity, we assume that the aggregate quality is additive and that all sensors provide measurements of equal quality at each time slot, which is taken to be 1. Therefore  $W(t)$  is analogous to the number of active sensors at time  $t$ , thus let  $W(t)$  denote the number of active sensors at time  $t$ .

Define a utility function  $U(W(t))$  that reflects the utility derived at each time slot  $t$  from the data submitted by the active sensors at time  $t$ . We can safely assume that  $U(\cdot)$  is a concave function, thus capturing the fact the higher the aggregate quality is, the less the additional increase in utility is per unit of additional quality i.e. by having additional activated sensors.

1) *Centralized control policy:* The global performance objective is the long-term average utility stemming out of the quality of collected measurements. First, we consider a centralized control policy, according to which a central platform controller schedules sensor activations so as to maximize this long-term average utility,

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T \mathbb{E}_t[U(W(t))] = \mathbb{E}[U(W(t))], \quad (11)$$

where the first expectation is with respect to probabilities  $\Pr[W(t) = m]$  for  $m = 1, 2, \dots$ , and the second expectation is with respect to the stationary distribution of  $W(t)$ ,  $\{w_0, w_1, w_2, \dots\}$ , with  $w_m = \lim_{t \rightarrow \infty} \Pr[W(t) = m]$ , for  $m = 1, 2, \dots$ .

**Conjecture 1.** *The policy that maximizes objective (11) is of threshold type.*

From Jensens inequality, for a random variable  $X$  and a concave function  $U(\cdot)$ , it is  $\mathbb{E}[U(X)] \leq U(\mathbb{E}[X])$ . Equality holds if and only if  $X = \mathbb{E}[X]$ , i.e. when random variable  $X$  is constant. In our setting, Jensen's inequality implies

$$\mathbb{E}[U(W(t))] \leq U(\mathbb{E}[W(t)]), \quad (12)$$

with equality if and only if  $W(t)$  is fixed across time and equal to  $\mathbb{E}[W(t)]$ . This means that the optimal policy should continuously maintain a balanced number of active sensors across time. A class of policies that intuitively achieve that goal is the threshold-based policies.

Consider the following threshold-based sensor activation policy. There exists a threshold number of active sensors,  $\bar{W}$ . At time  $t$  of arrival of a new sensor, the controller decides whether the sensor will be activated immediately or sometime later by the end of its availability duration. If the instantaneous number of active sensors  $W(t) \leq \bar{W}$ , the sensor is activated immediately, otherwise the sensor enters a queue of other queued sensors that all await activation. Queued sensors are activated at some point before the sensor leaves the system when the number of active sensors  $W(t)$  drops below  $\bar{W}$  after an active sensor leaves the system.

It would be interesting to further research the specifics of this and of other similar policies in terms of their properties and optimality.

2) *Individual sensor activation policy*: In section II, the decision of each sensor to be on or off at a time slot was guided by the time-varying costs and advertised prices, and it did not affect the decision of other sensors. Consider now the following slightly different setting. Arriving sensors are initially inactive, and each of them faces an optimal stopping problem with horizon  $1/s$ , during which it decides whether to become active or continue to be inactive. Assume that the price paid by the platform per time slot is common for all sensors, and that it may also be changing with time. One can easily apply the rationale of (3) to show that the optimal policy for each sensor is

$$\begin{aligned} &\text{activate, if } c_k - p_k \leq (T - k)(p - \mathbb{E}[C]) \\ &\text{stay inactive, if } c_k - p_k > (T - k)(p - \mathbb{E}[C]). \end{aligned} \quad (13)$$

Intuitively, this time-varying price can be set to be equal to

$$p(t) = \frac{dU(W(t))}{dW(t)}, \quad (14)$$

and it represents the differential utility that a sensor brings to the global objective if it becomes activated.

If the arrival process of sensors is Poisson, the aggregate quality i.e. the number of active sensors  $W(t)$  evolves as a Markov chain. It turns out by Little's theorem that  $\mathbb{E}[W(t)] = \lambda/s$ . However, the strategies of individual sensors determine the statistics of  $W(t)$ . When multiple sequential decision makers operate, the decision of each sensor affects the statistics of the number of active sensors  $W(t)$ , which in turn affects the optimal stopping decisions of sensors.

The first question is whether the individual sensor optimal stopping strategy is a Nash equilibrium. A Nash equilibrium is a sensor strategy such that no sensor may achieve a better net payoff by unilaterally changing its own strategy, assuming that all other sensors' strategies remain fixed.

We conjecture that the individual sensor optimal stopping strategies indeed constitute a Nash equilibrium. Assume that all sensors' strategies do not change. Then, the statistics of  $W(t)$ , namely the probabilities  $w_m = \lim_{t \rightarrow \infty} \Pr[W(t) = m]$  do not change, and therefore the statistics of the price process  $\{p(t)\}$  do not change as well. Therefore, the optimal policy (13) seems to be the best choice that a sensor can take given that other sensors do not change their strategies.

Another interesting issue would be to evaluate the global performance in terms of long-term expected utility that stems from individual sensor optimal-stopping strategies and to compare it with the optimal performance for a centralized optimal policy of sensor activation scheduling.

#### IV. CONCLUSION

In this work in progress, we studied the problem of sensor activation in mobile crowdsensing, towards contributing to a global utility stemming from quality of collected sensor data. We started by having as a basis the single-sensor activation problem which can be formulated as an optimal-stopping sequential decision one. We believe that bringing this sequential decision-making perspective for each sensor to the fore makes the mobile crowdsensing model more realistic.

We next outlined two interesting future directions for research. The first one seeks to investigate incentive payment allocation to sensors with the objective to maximize average quality of collected data. A static setting is considered, where the probability that a sensor is active at a time slot is computed based on the optimal stopping policy of each sensor, and these probabilities are then inserted in the optimization problem.

The second setting is a dynamic one and calls for dynamic centralized sensor activation policies in order to maximize the long-term average utility stemming from quality of collected data. To this end, characterizing equilibrium rational policies of sensors and assessing the impact of these policies on the global performance objective are directions that warrant further investigation.

#### REFERENCES

- [1] R. K. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: Current state and future challenges", *IEEE Comm. Mag.*, vol.49, no.11, pp.32-39, Nov. 2011.
- [2] I. Koutsopoulos, "Optimal incentive-driven design of participatory sensing systems", in *Proc. IEEE INFOCOM*, 2013.
- [3] L. Duan, T. Kubo, K. Sugiyama, J. Huang, T. Hasegawa and J. Walrand, "Incentive mechanisms for smartphone collaboration in data acquisition and distributed computing", in *Proc. IEEE INFOCOM*, 2012.
- [4] D. Yang, G. Xue, X. Fang, and J. Tang, "Crowdsourcing to smartphones: Incentive mechanism design for mobile phone sensing", in *Proc. ACM MobiCom*, 2012.
- [5] L. Gao, F. Hou, and J. Huang, "Providing long-term participation incentive in participatory sensing", in *Proc. IEEE INFOCOM*, 2015.
- [6] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing and X. Mao, "Incentives for mobile crowdsensing: A survey", *IEEE Comm. Surveys and Tut.*, vol.18, no.1, pp.54-67, 1st Quart. 2016.
- [7] M. Karaliopoulos, I. Koutsopoulos and M. Titsias, "First Learn then Earn: Optimizing Mobile Crowdsensing Campaigns through Data-driven User Profiling", in *Proc. ACM MobiHoc*, 2016.
- [8] Y. Ueyama, M. Tamai, Y. Arakawa, and K. Yasumoto, "Gamification-based incentive mechanism for participatory sensing", in *Proc. IEEE PERCOM Workshops*, 2014.
- [9] K. Szajowski, "Markov stopping games with random priority", *Mathem. l Methods for Oper. Research*, vol. 39, no.1, pp. 69-84, Feb. 1994.
- [10] X. He, H. Dai, P. Ning and R. Dutta, "A Multi-player Markov Stopping Game for Delay-tolerant and Opportunistic Resource Sharing Networks", in *Proc. IEEE INFOCOM*, 2016.
- [11] P. Yang, K. Iyer, and P. Frazier, "Mean Field Equilibria for Competitive Exploration in Resource Sharing Settings", in *Proc. Int. Conf. on World Wide Web (WWW)*, 2016.
- [12] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, vol.1, Athena Scientific, 1995.