

# Energy Efficient Data Collection in Opportunistic Mobile Crowdsensing Architectures for Smart Cities

Andrea Capponi<sup>†</sup>, Claudio Fiandrino<sup>‡†</sup>, Dzmitry Kliazovich<sup>♦</sup>, Pascal Bouvry<sup>†</sup>

<sup>†</sup> Computer Science and Communications Research Unit, University of Luxembourg, Luxembourg

<sup>‡</sup> Imdea Networks Institute, Leganés, Spain

<sup>♦</sup> ExaMotive, Luxembourg

E-mails: <sup>†</sup>{firstname.lastname}@uni.lu, <sup>‡</sup>claudio.fiandrino@imdea.org, <sup>♦</sup>kliazovich@ieee.org,

**Abstract**—Smart cities employ latest information and communication technologies to enhance services for citizens. Sensing is essential to monitor current status of infrastructures and the environment. In Mobile Crowdsensing (MCS), citizens participate in the sensing process contributing data with their mobile devices such as smartphones, tablets and wearables. To be effective, MCS systems require a large number of users to contribute data. While several studies focus on developing efficient incentive mechanisms to foster user participation, data collection policies still require investigation. In this paper, we propose a novel distributed and energy-efficient framework for data collection in opportunistic MCS architectures. Opportunistic sensing systems require minimal intervention from the user side as sensing decisions are application- or device-driven. The proposed framework minimizes the cost of both sensing and reporting, while maximizing the utility of data collection and, as a result, the quality of contributed information. We evaluate performance of the framework with simulations, performed in a real urban environment and with a large number of participants. The simulation results verify cost-effectiveness of the framework and assess efficiency of the data generation process.

## I. INTRODUCTION

Cities are facing the challenge for a sustainable and efficient development. While half of the worldwide population lives in metropolitan areas, this figure is projected to increase significantly in the next three decades [1]. Sustainable development requires precise monitoring of resource utilization, including water, electricity, gas among the others. Sensing plays an essential role on this regard. Sensors enable monitoring of the current status of infrastructures, transportation systems, environment and health [2]. Deploying new sensing infrastructures is typically expensive, while active participation of citizens can improve spatial coverage of already deployed infrastructures with no need of further investments. Mobile devices include a rich set of built-in sensors, suitable for monitoring in multiple domains, such as smart homes, personal health care, public safety [3]. Mobility and intelligence of human participants guarantee higher coverage and better context awareness, if compared to traditional sensor networks. In addition, users maintain by themselves the mobile devices and provide periodic recharge.

Mobile Crowdsensing (MCS) leverages active participation of citizens<sup>1</sup> in sensing activities, which are required to

contribute data from sensors of their mobile devices. This contribution is reported using commonly available cellular or WiFi communication technologies to a collector, typically located in the cloud, for data processing and analysis. In MCS, data collection frameworks define the set of steps necessary to collect and report the information. Frameworks can be *participatory* or *opportunistic* [4]. In participatory sensing systems, users are directly involved in the sensing process and they need to accomplish specified tasks (e.g., taking a picture). On the contrary, opportunistic sensing systems minimize user involvement and sensing decisions are device-driven (e.g., recording ambient light). In both types of frameworks, proper recruitment of the participants is essential [5]. Hassani et al. [6] proposed Context-Aware Task Allocation (CATA), which allocates tasks after having recruited users in opportunistic MCS systems. The recruitment policy selects the most appropriate users to fulfill the sensing task by determining the similarity between the participants and the tasks. The policy considers energy consumption of sensing and data delivery operations to determine the eligibility of the participants. Zhu et al. [7] focus on the upload planning of real-time information collected by IoT devices, where the network context changes dynamically. They present a prototype system called SCALECycle, and propose a two-phase approach exploiting the knowledge of IoT deployments and heterogeneous nature of IoT applications.

Devising efficient frameworks for data collection is fundamental. MCS follows a Sensing as a Service (S<sup>2</sup>aaS) business model, which makes data collected from sensors available to cloud users [8]. Consequently, companies and organizations have no longer the need to acquire an infrastructure to perform a sensing campaign, but they can exploit existing ones in a pay-as-you-go basis. Efficiency of S<sup>2</sup>aaS models is defined in terms of the revenues obtained and the costs. The organizer of a sensing campaign, such as a government agency, an academic institution or a business corporation, sustains costs to recruit and compensate users for their involvement [9]. The users sustain costs while contributing data too, i.e., the energy spent from the batteries for sensing and reporting data and, eventually, the data subscription plan if cellular connectivity is used for reporting.

In this paper, we propose a distributed framework for data collection in opportunistic MCS systems. The framework aims to minimize the cost of sensing and reporting by estimating in distributed manner the *smartphone sensing potential*, i.e., the utility of performing such operations. At the same time,

Dr. Claudio Fiandrino developed this work as a Ph.D. student at the University of Luxembourg.

<sup>1</sup>In the rest of the paper, we will use the terms citizens, participants and users interchangeably.

the quality of contributed information for the system is guaranteed through *data collection utility* feedback. The collector announces estimated utilities by periodically broadcasting the type and the minimum amount of data required to capture a physical phenomena with a given level of accuracy. To maximize the value of the collected data to the system at the minimum cost of sensing, the crowd participants sense and report data when there is a match between the sensing potential and the data collection utility feedback the collector provides. A match ensures that the mobile devices sustain a cost to produce useful data for the cloud collector. On one hand, the mechanism prevents the users to contribute too much, e.g., draining the user battery completely for continuous use, or too little. On the other hand, being the framework completely distributed, each device is responsible to decide the timing and the duration of the contribution. This minimizes manual intervention of the users, lowering the barriers for participation to the sensing process.

## II. THE OPPORTUNISTIC DATA COLLECTION FRAMEWORK

The mobile devices decide to perform sensing and reporting independently one from another one. For each sensor  $s$ , the decisions occur on the basis of the *data collection utility*  $d_s^a$ , the *smartphone sensing potential*  $sp_s$ , the *environmental context* of the device  $C_s$  and a threshold  $\delta$ , which defines the amount of contribution each participant provides. Consequently, in this work we do not consider correlation among the sensors for taking decisions. The framework operates in distributed fashion because each device is responsible to locally compute all the parameters necessary to determine whether to participate to the sensing process with the sole exception of  $d_s^a$ . The derivation of the parameters  $d_s^a$ ,  $sp_s$ ,  $C_s$  and  $\delta$  is illustrated in more detail in the following paragraphs.

The decision policy is formally defined as follows:

$$C_s \cdot [\gamma \cdot sp_s + (1 - \gamma) \cdot d_s^a] > \delta, \quad (1)$$

where  $\gamma$  is a balancing coefficient which assumes real values in the range  $[0, 1]$ . The parameter  $\gamma$  is computed by the collector and broadcasted to the participants. High values of  $\gamma$  give more relevance to the *smartphone sensing potential*, while low values of  $\gamma$  make the *data collection utility* term dominant. When the contribution given by the first member of (1) is above  $\delta$ , the smartphone performs both sensing and reporting. When the environmental context is unfavorable ( $C_s = 0$ ), the devices never perform sensing and reporting. Unfavorable occurs, when, for example, the smartphone is in the pocket and the device needs to record current ambient light. The design choice prevents the devices of performing useless operations and allows to conserve resources, assuming the utility of such samples would be low for the collector.

### A. Data Collection Utility

Following the Sensing-as-a-Service (S<sup>2</sup>aaS) paradigm, the collector is located in the cloud. Having received a request for sensing in a given area, the cloud informs the sensors and the mobile devices located in that area. They accomplish the request and collect the information. Based on the sensing

interest, which is assumed to be known in advance, cloud collector decides which samples need to be collected. Requests can come from different applications and may require samples from different location areas and types of sensors.

To define *data collection utility*, the cloud collector partitions the monitored region in a set of areas  $\mathcal{A}$ . In each area  $a \in \mathcal{A}$ , the mobile devices generate samples from a different set of sensors  $\mathcal{S}$ . We describe the average number of samples  $\overline{N}_s^a|_t$  generated from sensor  $s$  in area  $a$  during the timeslot  $t$  through the Exponential Weighted Moving Average filter (EWMA):

$$\overline{N}_s^a|_t = \sigma \cdot N_s^a|_t + (1 - \sigma) \cdot \overline{N}_{s-1}^a|_t, \quad (2)$$

where  $N_s^a|_t$  corresponds to the number of samples collected from sensor  $s$  in timeslot  $t$  in area  $a$  and  $\overline{N}_{s-1}^a|_t$  is its previous value. The parameter  $\sigma$  is the exponential weighting coefficient. High values of  $\sigma$  limit the contribution of older values whose utility, from the collector point of view, is lower than the contribution of newly generated samples.

High values of  $\overline{N}_s^a|_t$  indicate a large number of samples have been already received at the cloud collector and further reporting is not needed. Viceversa, low values of  $\overline{N}_s^a|_t$  indicate the need for more samples and the data collection utility is high. These considerations suggest that the data collection utility can be defined as the following sigmoid function:

$$d_s^a = \frac{1}{1 + e^{-\frac{\varphi_s}{\rho_s} \cdot (-\overline{N}_s^a|_t + (1 - \frac{\rho_s}{2}))}}, \quad (3)$$

where  $\varphi_s$  and  $1 - \rho_s/2$  coefficients control position and the speed of the incline. The data collection utility can assume real values in the range  $[0, 1]$ .

The cloud collector computes  $d_s^a$  per area  $a$  and sensor  $s$  and informs the participants in each area by transmitting periodically beacon messages.

### B. Smartphone Sensing Potential

Each smartphone computes individually and independently from the others the *smartphone sensing potential* as function of locally spent energy  $E_s$  for sensing and reporting. The approach is similar to the methodology proposed in [10], where the goal is to minimize prediction error for MCS tasks with classification purposes. Similarly to the data collection utility, the relation between  $sp_s$  and  $E_s$  is defined with a sigmoid function:

$$sp_s = \frac{1}{1 + e^{-\frac{\zeta_s}{\theta_s} \cdot (-E_s + (1 - \frac{\theta_s}{2}))}}. \quad (4)$$

The smartphone sensing potential  $sp_s$  can assume real values in the range  $[0, 1]$ , while the parameters  $1 - \theta_s/2$  and  $\zeta_s$  control position of the center and speed of the incline respectively. On one hand, a steep increase of the incline reduces transition between low and high values of the smartphone sensing potential and makes the device to work preferentially in on/off mode, i.e., the device is either an excellent or the worst candidate for contribution. On the other hand, the position of the center of the function defines which range of values of  $E_s$  makes the device to be an excellent candidate for contribution.

The energy the mobile devices spend in contributing data can be attributed to sensing ( $E_s^c$ ) and reporting ( $E_s^r$ ) operations:

$$E_s = E_s^c + E_s^r. \quad (5)$$

For sensing, the contribution  $E_s^c$  has to be taken into account only if sensor  $s$  is not already in use by another application, which is defined by the *utilization context* of  $s$ ,  $U_s$ . Otherwise, the energy spent by  $s$  is equal to zero.  $E_s^c$  is defined as follows:

$$E_s^c = \bar{E}_s^c \cdot U_s, \quad (6)$$

where  $\bar{E}_s^c$  is the actual energy spent by sensor  $s$  and the utilization context  $U_s$  is defined as:

$$U_s = \begin{cases} 0, & \text{if the sensor } s \text{ is used by another application;} \\ 1, & \text{otherwise.} \end{cases} \quad (7)$$

Data reporting implies delivering information collected from the set of sensors  $\mathcal{S}$  and transmitting it to the cloud collector wirelessly. Reporting is always performed at the beginning of the timeslot  $t$  for samples collected during timeslot  $t-1$ . Energy cost related to communication  $E_s^r$  depends on the employed technology, LTE or WiFi, and is defined as follows:

$$E_s^r = \begin{cases} E^W, & \text{if WiFi or both WiFi and LTE are enabled;} \\ E^{L1}, & \text{if WiFi is disabled and LTE is idle state;} \\ E^{L2}, & \text{if WiFi is disabled and LTE is connected state.} \end{cases} \quad (8)$$

When both WiFi and LTE interfaces are active, transmissions take place via WiFi as it is more energy efficient [11], [12] and users do not consume the data plan they pay to the cellular operators [13]. The energy  $E^W$  spent during the transmission time  $\tau_{tx}$  is defined as:

$$E^W = \int_0^{\tau_{tx}} P_{tx}^W dt, \quad (9)$$

where  $P_{tx}^W$  is the power consumed for transmissions of WiFi packets generated at rate  $\lambda_g$  [14]:

$$P_{tx}^W = \rho_{id} + \rho_{tx} \cdot \tau_{tx} + \gamma_{xg} \cdot \lambda_g. \quad (10)$$

The parameters  $\rho_{id}$ ,  $\rho_{tx}$  and  $\gamma_{xg}$  represent the energy in idle mode, the transmission power and the energy cost to elaborate a generated packet.

The Radio Resource Control (RRC) state machine and the simplified model proposed in [15] are used to model LTE power consumption. The model defines different energy consumption levels in relation to the initial state. Although initial states of the system can be *connected*, *tail* and *idle* [11], we focus on *connected* and *idle* states, as *idle* state can be considered as a worst case scenario of *tail* state.

Whenever the smartphone is idle and needs to communicate, it transitions into the *connected* state and after the transmission is over it goes into the *tail* state before finally returning back to *idle*. In this case, the energy consumption for the smartphone during reporting can be defined as:

$$E^{L1} = P_p \cdot T_p + P_{tx}^L \cdot T_{tx} + P_{tx}^L \cdot DRX_{IT} + P_{DRX} \cdot RRC_{IT}, \quad (11)$$

where  $T_p$  and  $P_p$  are the promotion delay and power,  $T_{tx}$  and  $P_{tx}$  are time and power transmission,  $DRX_{IT}$  is the Discontinuous Reception (DRX) Inactivity Timer,  $P_{DRX}$  is the power consumed when the smartphone is in one of the two DRX modes and  $RRC_{IT}$  is the RRC Inactivity Timer.

When the smartphone is already in RRC connected state and transmitting, its energy consumption can be defined considering

only the contribution of signal transmission:

$$E^{L2} = \int_0^{T_{tx}} P_{tx}^L dt. \quad (12)$$

The power consumption for transmitting data  $P_{tx}$  is given by the model of [11]:

$$P_{tx}^L = \alpha_{ul} \cdot T_{ul} + \beta, \quad (13)$$

where  $T_{ul}$  represents the uplink throughput and the parameters  $\alpha_{ul}$  and  $\beta$  are the power spent during transmission and the base power respectively [11].

### C. Profiling the Environmental Context

Environmental context  $C_s$  of sensor  $s$  defines the set of facts and circumstances happening around the mobile device, such as the location or the mobility pattern [16]. Having knowledge of the environmental context is essential to avoid performing sensing under unfavorable conditions that affect the overall energy budget without providing any benefit for the collector. The utility of each sample is defined according to the environmental context. Although a few practical solutions exist [17], estimating the environmental context is not simple. For the sake of simplicity, in this work we assume  $C_s$  to take binary values:

$$C_s = \begin{cases} 1, & \text{if the sample contributes to the sensing objective;} \\ 0, & \text{otherwise.} \end{cases} \quad (14)$$

When starting a new sensing campaign, the collector first broadcasts profiles defining the environmental context to all the devices. This is because each sensing campaign is application-based and environmental context profiles must be tailored to it.

### D. Threshold for Sensing and Reporting

Sensing and reporting operations occur when *data collection utility* and *smartphone sensing potential* are greater than a threshold  $\delta$ , which means that the mobile devices sustain a cost to produce useful data for the cloud collector. The mechanism prevents the users from contributing too much or too few of data. The threshold  $\delta$  is computed locally at the mobile device. A proper setting of this parameter is essential to define the amount of data each device opportunistically contributes. To define  $\delta$ , we take into account current *level of battery* of the devices (denoted as  $B$ ) and the *amount of reported data* that devices have already contributed to the system (denoted as  $D$ ). The parameter  $\delta$  is defined as follows:

$$\delta = f(\delta_b, \delta_d), \quad (15)$$

where all the parameters  $\delta$ ,  $\delta_b$  and  $\delta_d$  are real values in the range  $[0, 1]$ . Both  $\delta_b$  and  $\delta_d$  are function of  $B$  and  $D$  respectively. When any of the two parameters assumes a value equal to 1,  $\delta$  becomes 1 as well. As a result, the devices stop contributing upon meeting any of the two conditions: a low remaining charge of the battery or a high amount of data already contributed. Although  $f$  can be chosen arbitrarily, in this work we provide equal weights to  $\delta_b$  and  $\delta_d$ :

$$f(\delta_b, \delta_d) = (\delta_b + \delta_d)/2. \quad (16)$$

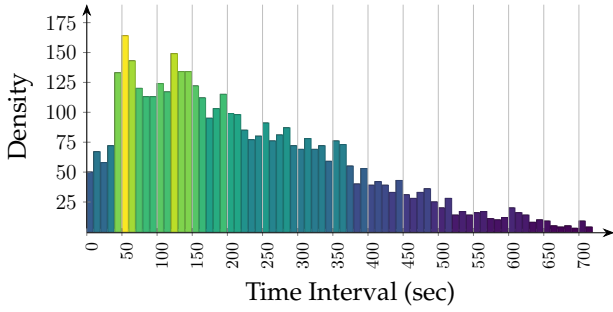


Fig. 1. User distribution of mobility trace “kth/walkers”

**Level of Battery:** We define the *level of battery*  $B$  as the remaining charge of the device, where  $0 \leq B \leq 1$ . High values of  $B$  correspond to a high level of charge and the device can contribute data from its sensors. On the other hand, when the battery is almost empty the users would like to preserve the remaining charge. These considerations suggest that the relation between  $B$  and  $\delta_b$  follows a negative exponential law. As a result, low values of  $B$  will make  $\delta_b$  to assume values close to 1.  $\delta_b$  is defined as follows:

$$\delta_b = \alpha^{\lambda \cdot B}. \quad (17)$$

The parameter  $\alpha$  can assume arbitrary real values between  $[0, 1]$ , and  $\lambda > 1$ .

**Amount of Reported Data:** The *amount of reported data*  $D$  defines a total amount of data that is delivered using WiFi ( $D_s^W$ ) and cellular ( $D_s^L$ ) connections and is contributed by the set of sensors  $\mathcal{S}$  of a single device.  $D$  is defined as follows:

$$D = \sum_{s \in \mathcal{S}} D_s^W + D_s^L. \quad (18)$$

The more data the device contributes, the higher the values parameter  $\delta_d$  assumes. As a result, the devices that have already delivered a significant amount of data to the collector, make further contribution only if it is necessary. On the other hand, low values of the parameter  $\delta_d$  will facilitate contribution from the devices that have provided little contribution to the system. The relation between  $D$  and  $\delta_d$  is modeled as follows:

$$\delta_d = \begin{cases} 1, & \text{if } D \geq D_{\max}; \\ \log\left(1 + \frac{D}{D_{\max}}\right), & \text{otherwise;} \end{cases} \quad (19)$$

where  $D_{\max}$  is the maximum amount of data each device is willing to deliver. This parameter can be tuned by the users periodically.

### III. PERFORMANCE EVALUATION

We evaluate performance of the proposed framework with a custom crowdsensing simulator, CrowdSenSim [18]. We first present its detailed functionality, then the simulation scenario and the obtained results.

#### A. Simulation Set-up

CrowdSenSim supports pedestrian mobility in citywide scenarios and users contribute data according to the designed opportunistic framework. Data generation exploits sensors commonly available in mobile devices. The simulation results can be obtained at the level of individual devices as well as the

system level, which helps to analyze data reporting progress and efficiency of the employed crowdsensing techniques.

The layout of the city consists of a set of coordinates  $C$  containing information on  $\langle \text{latitude}, \text{longitude}, \text{altitude} \rangle$  defining the streets of the city where the users move.  $C$  can be obtained with online tools like OpenStreetMaps or DigiPoint. In this work, we exploit DigiPoint, which is a crowdsourced application providing free access to street-level maps<sup>2</sup>. The center of Luxembourg city covers an area of 1.11 km<sup>2</sup> with a population of 110 499 inhabitants as of end of 2015. The city center of Trento occupies an area of 1.18 km<sup>2</sup> and has a population 117 317 inhabitants as of beginning of 2016 while the city center of Madrid covers approximately an area of 5.23 km<sup>2</sup> with a resident population of 149 718 residing inhabitants. The number of participants is fixed to 20 000.

Each participant has only one mobile device and walks for a period of time that is uniformly distributed between  $[10, 20]$  minutes with an average speed uniformly distributed between  $[1, 1.5]$  m/s. We employ two user arrival patterns. In the first, the users arrive with a uniform probability between 8:00 AM - 1:40 PM. The second arrival pattern is based on real-world traces, which are the results of a study on pedestrian mobility and are public available on Crowdad (ostermalm\_dense\_run2) [19]. Fig. 1 shows the probability density function of the user arrival resulting from the study of the traces. In practice, to obtain the results presented later, the density computed in Fig. 1 was scaled and adapted to the considered arrival time period.

Data generation takes place using as sensing equipment the FXOS8700CQ 3axis linear accelerometer from Freescale Semiconductor<sup>3</sup> and the BMP280 from Bosch<sup>4</sup>, which is a digital pressure and temperature sensor. Communications occur over the WiFi link, having obtained the precise location of WiFi hotspots in form of  $\langle \text{latitude}, \text{longitude} \rangle$ . Table I presents the detailed information on communication and the parameters. The parameter  $\delta_b$  is determined using (17) by setting  $\alpha = 0.7$  and  $\lambda = 10$ . These values proved by the conducted analytical analysis to perform better, which we omit for the space reasons. The parameter  $\delta_d$  is set to 0 during initialization. The utilization context  $U_s$  is modelled using random uniformly distribution in the range  $[0, 1]$  and is generated during each timeslot for each user device. For the analysis, the parameter  $\gamma$  is fixed to 1/2 to give equal importance to the data collection utility and the smartphone sensing potential.

#### B. Simulation results

Fig. 2 presents the distribution of users and their energy spent for sensing with the uniform- and traces-based user arrival patterns. For demonstration purposes, we show the results obtained for the sole city of Luxembourg. As expected, the user arrival pattern does not influence consumption of energy, which only depends on the amount of time the users generate data. The profiles of Fig. 2(b) and Fig. 2(a) follow a normal distribution as the users contribute data for time periods as low

<sup>2</sup>DigiPoint 3, <http://www.zonums.com/gmaps/digipoint.php>

<sup>3</sup><http://www.nxp.com/files-static/sensors/doc/data-sheet/FXOS8700CQ.pdf>

<sup>4</sup><https://www.bosch-sensortec.com/bst/products/all-products/bmp280>

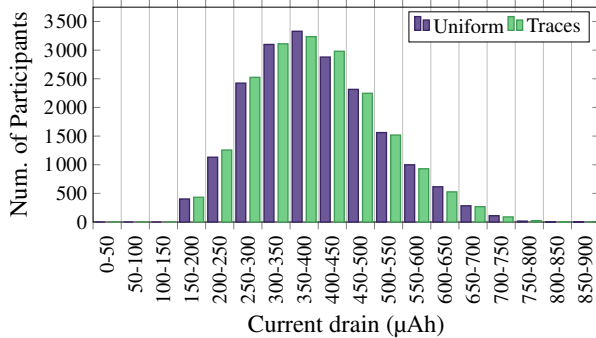
TABLE I  
SENSOR AND COMMUNICATION EQUIPMENT PARAMETERS USED FOR PERFORMANCE EVALUATION

SENSOR	PARAMETER	VALUE	UNIT
Accelerometer	Sample rate	50	Hz
	Sample size	12	Bits
	Current	35	$\mu\text{A}$
Temperature	Sample rate	182	Hz
	Sample size	16	Bits
	Current	182	$\mu\text{A}$
Pressure	Sample rate	157	Hz
	Sample size	16	Bits
	Current	423.9	$\mu\text{A}$

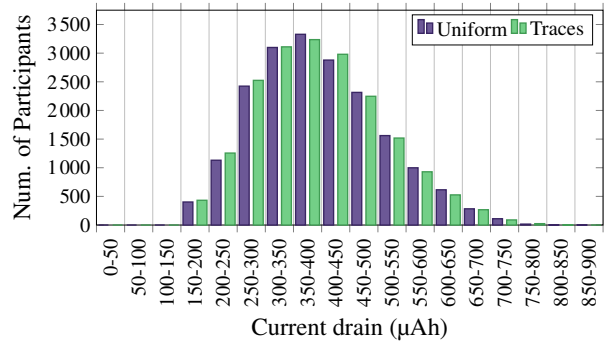
(a) Sensor Equipment

SYMBOL	VALUE	UNIT	DESCRIPTION
$\rho_{id}$	3.68	W	Power in idle mode
$\rho_{tx}$	0.37	W	Transmission power
$\rho_{rx}$	0.31	W	Reception power
$\lambda_g$	1000	fps	Rate of generation of packets
$\gamma_{xg}$	$0.11 \cdot 10^{-3}$	J	Energy cost to elaborate a generated packet

(b) Communication Equipment



(a) Sensing Cost



(b) Communication Cost

Fig. 2. Energy spent for sensing and communication

as 10 minutes up to a maximum of 20 minutes. Current drain of sensing operations is on average  $373.41 \mu\text{Ah}$  and  $368.80 \mu\text{Ah}$  for uniform- and traces-based arrival patterns respectively. In the worst case, few users experience a cost that is nearly more than double with respect to the average. Comparing to the battery capacity available in modern smartphones, which is in the order of 2000 mAh, it is possible to conclude that the energy cost for sensing is negligible with respect to the energy spent for communications (see Fig. 2(b)).

Fig. 3 shows the geographical distribution of the amount of collected data at the end of the simulation period for Luxembourg, Trento and Madrid. To better analyze the data generation process, we define a new metric called Sample Distribution (SD). SD measures the amount of generated samples per meter and is defined as follows:

$$\text{SD} = N_s^a | \bar{\Delta}, \quad (20)$$

where  $\bar{\Delta}$  is the average distance between samples and  $N_s^a |$  is the number of samples generated per sensor  $s$ , area  $a$  at timeslot  $t$ . The parameter  $\bar{\Delta}$  is defined as follows:

$$\bar{\Delta} = \frac{\sum_{i,j}^{i \geq j} d(i,j)}{n(n-1)}. \quad (21)$$

The term  $d(i, j)$  is the distance (in meters) between the location where the samples  $i$  and  $j$  were generated and the denominator accounts for the number of pairs of samples.

Fig. 4 shows the distribution of SD for Luxembourg for the entire simulation period. The city has been divided into 5 areas of similar size. In this experiment, the users are located with the uniform arrival pattern. It is interesting to notice that

the lowest values of SD occur for the initial and final time intervals (8:00 AM - 9:00 AM and 1:00 PM -2:00 PM). During the initial and final time intervals the number of participants is lower than in the other intervals as the simulator locates the users with a uniform distribution between 8:00 AM and 1:40 PM and they move for at maximum 20 minutes. In this analysis, only accelerometer samples were utilized. The SD metric weakly depends on the size of the area. Although being a bit wider than Area 3, a large part of Area 2 is a public park with a fewer number of streets. As the SD metric measures the distribution of the samples taking into account the location where they have been generated, high density areas, such as Area 5, exhibit high values of SD.

#### IV. CONCLUSION

This paper proposes a new data collection framework for opportunistic MCS systems. The framework works in distributed fashion and aims at minimizing the cost of sensing and reporting of data for the participants, while maximizing data collection utility. The performance of the framework is verified through simulations in realistic urban environments. We analyze the relation between the cost imposed on participants and the efficiency of the data generation process. Results highlight that the major contribution to energy consumption is attributed to reporting, not sensing. The simulation results confirm effectiveness of the proposed approach for a large number of participants. As future work, we plan to implement the current model with an application to verify the performance of the framework experimentally.



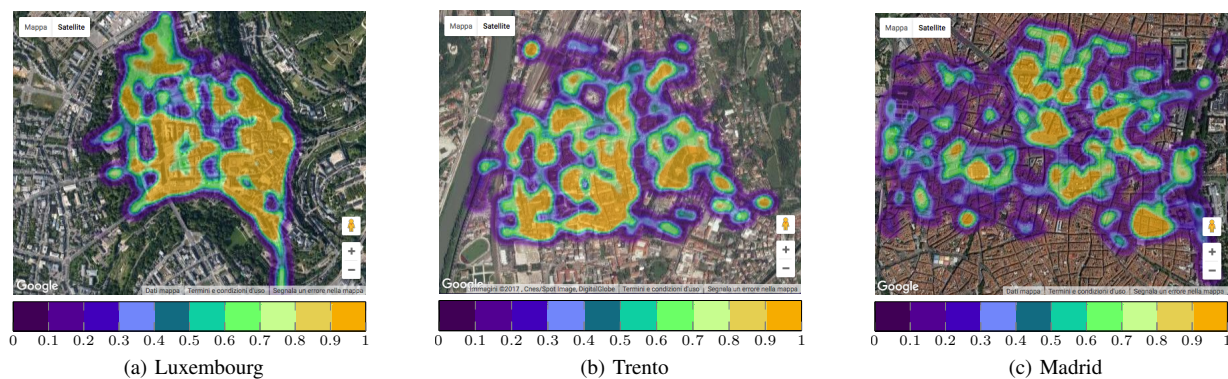


Fig. 3. Normalized distribution of amount of collected data for the different cities over the time period 8:00 AM - 2:00 PM

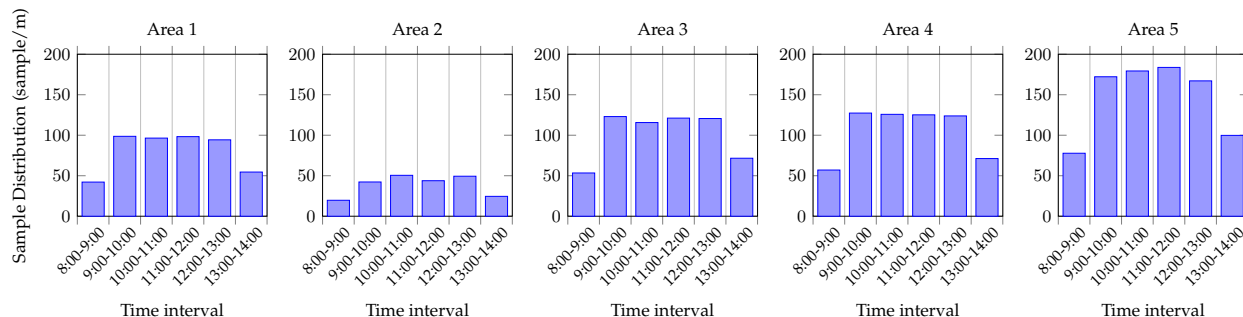


Fig. 4. Sample distribution (SD) for the different areas over the time period 8:00-14:00

## REFERENCES

- [1] N. B. Grimm, S. H. Faeth, N. E. Golubiewski, C. L. Redman, J. Wu, X. Bai, and J. M. Briggs, "Global change and the ecology of cities," in *Science*, vol. 319, no. 5864, 2008, pp. 756–760.
- [2] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things Journal*, vol. 1, no. 1, pp. 22–32, 2014.
- [3] C. Perera, A. Zaslavsky, P. Christen, and D. Georgakopoulos, "Sensing as a service model for smart cities supported by Internet of Things," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 1, pp. 81–93, 2014.
- [4] W. Khan, Y. Xiang, M. Aalsalem, and Q. Arshad, "Mobile phone sensing systems: A survey," *IEEE Communications Surveys Tutorials*, vol. 15, no. 1, pp. 402–427, First 2013.
- [5] C. Fiandrino, B. Kantarci, F. Anjomshoa, D. Kliazovich, P. Bouvry, and J. Matthews, "Sociability-driven user recruitment in mobile crowdsensing internet of things platforms," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.
- [6] A. Hassani, P. D. Haghighi, and P. P. Jayaraman, "Context-aware recruitment scheme for opportunistic mobile crowdsensing," in *IEEE ICPADS*, Dec 2015, pp. 266–273.
- [7] Q. Zhu, M. Y. S. Uddin, Z. Qin, and N. Venkatasubramanian, "Upload planning for mobile data collection in smart community Internet-of-Things deployments," in *IEEE SMARTCOMP*, 2016, pp. 1–8.
- [8] A. Capponi, C. Fiandrino, C. Franck, U. Sorger, D. Kliazovich, and P. Bouvry, "Assessing performance of internet of things-based mobile crowdsensing systems for sensing as a service applications in smart cities," in *IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Dec 2016, pp. 456–459.
- [9] K. Han, C. Zhang, and J. Luo, "Taming the uncertainty: Budget limited robust crowdsensing through online learning," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1462–1475, June 2016.
- [10] X. Qiang and Z. Rong, "When data acquisition meets data analytics: A distributed active learning framework for optimal budgeted mobile crowdsensing," in *IEEE International Conference on Computer Communications (INFOCOM)*, May 2017.
- [11] J. Huang, F. Qian, A. Gerber, Z. M. Mao, S. Sen, and O. Spatscheck, "A close examination of performance and power characteristics of 4G LTE networks," in *10th ACM International Conference on Mobile Systems, Applications, and Services*, ser. *MobiSys*, 2012, pp. 225–238.
- [12] C. Ragona, F. Granelli, C. Fiandrino, D. Kliazovich, and P. Bouvry, "Energy-efficient computation offloading for wearable devices and smartphones in mobile cloud computing," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2015, pp. 1–6.
- [13] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Y. Zomaya, "Network-assisted offloading for mobile cloud applications," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 5833–5838.
- [14] P. Serrano, A. Garcia-Saavedra, G. Bianchi, A. Banchs, and A. Azcorra, "Per-frame energy consumption in 802.11 devices and its implication on modeling and design," *IEEE/ACM Transactions on Networking*, vol. 23, no. 4, pp. 1243–1256, Aug 2015.
- [15] G. Foddiss, R. Garroppo, S. Giordano, G. Proccisi, S. Roma, and S. Topazzi, "LTE traffic analysis for signalling load and energy consumption trade-off in mobile networks," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 6005–6010.
- [16] O. Yurur, C. Liu, Z. Sheng, V. Leung, W. Moreno, and K. Leung, "Context-awareness for mobile sensing: A survey and future directions," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 68–93, Firstquarter 2016.
- [17] A. Rahmati, C. Shepard, C. Tossell, L. Zhong, and P. Kortum, "Practical context awareness: Measuring and utilizing the context dependency of mobile usage," *IEEE Transactions on Mobile Computing*, vol. 14, no. 9, pp. 1932–1946, Sept 2015.
- [18] C. Fiandrino, A. Capponi, G. Cacciatore, D. Kliazovich, U. Sorger, P. Bouvry, B. Kantarci, F. Granelli, and S. Giordano, "Crowdsensim: a simulation platform for mobile crowdsensing in realistic urban environments," *IEEE Access*, vol. PP, no. 99, pp. 1–1, February 2017.
- [19] S. T. Kouyoumdjieva, Ólafur Ragnar Helgason, and G. Karlsson, "CRAWDAD dataset kth/walkers (v. 2014-05-05)," Downloaded from <http://crawdad.org/kth/walkers/20140505>, May 2014.