This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

A Cost-Effective Distributed Framework for Data Collection in Cloud-based Mobile Crowd Sensing Architectures

Andrea Capponi, *Student Member, IEEE,* Claudio Fiandrino, *Student Member, IEEE,* Dzmitry Kliazovich, *Senior Member, IEEE,* Pascal Bouvry, *Member, IEEE,* and Stefano Giordano, *Senior Member, IEEE*

Abstract—Mobile crowd sensing received significant attention in the recent years and has become a popular paradigm for sensing. It operates relying on the rich set of built-in sensors equipped in mobile devices, such as smartphones, tablets and wearable devices. To be effective, mobile crowd sensing 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 sustainable framework for gathering information in cloud-based mobile crowd sensing systems with opportunistic reporting. The proposed framework minimizes cost of both sensing and reporting, while maximizing the utility of data collection and, as a result, the quality of contributed information. Analytical and simulation results provide performance evaluation for the proposed framework by providing a fine-grained analysis on the energy consumed. The simulations, performed in a real urban environment and with a large number of participants, aim at verifying the performance and scalability of the proposed approach on a large scale under different user arrival patterns.

Index Terms—Mobile crowd sensing, energy-efficient data collection, opportunistic sensing.

1 Introduction

Mobile crowd sensing (MCS) has become in the recent years an appealing paradigm for sensing and collecting data. In MCS, users contribute data gathered from sensors embedded in mobile devices such as smartphones, tablets and Internet of Things (IoT) devices like wearables. The information is then delivered to a collector, usually located in the cloud [1], [2]. The ubiquitous diffusion of mobile devices and the rich set of built-in sensors they are equipped with, have been the two main enablers leading to the success of MCS paradigm. Mobile devices have become essential for our everyday activities, including business, health-care and wellbeing, social activities and entertainment [3], [4]. Accelerometer, GPS, camera and microphone are only a representative set of sensors commonly available in mobile and IoT devices. A number of MCS-based applications builds on multi-sensing capabilities and finds applicability in different scenarios, including health-care, environmental monitoring, public safety and intelligent transportation systems such as traffic monitoring and management [5], [6], [7]. All these applications suit urban scenarios very well. Indeed, MCS is projected to be a key solution for smart cities, which aim at using ICT solutions to improve management and quality of everyday life of their citizens [8], [9].

The process of data collection includes a set of steps necessary to produce data and deliver it to the collector. In MCS systems, data collection can be either *opportunistic* or *participatory* [5], [6]. The user involvement in *opportunistic* sensing systems is minimal or none, which means that the decisions to perform sensing and report data are application- or device-driven. On the contrary, *participatory* sensing systems rely on active user engagement in the sensing process. For example, when users can spontaneously decide to contribute to the system after having received a specific task. Participatory sensing is tailored to crowd sensing architectures with a "central intelligence" responsible to assign tasks to the users, e.g., to ask one user to record a video in a given area at a given time. Unlike opportunistic sensing systems, the participatory paradigm imposes a higher cost on the user in terms of cooperation effort. Having devices or applications responsible for sensing lowers the burden for user participation and makes opportunistic sensing ideal for distributed solutions.

Devising efficient frameworks for data collection is fundamental. MCS follows a Sensing as a Service (S^2aaS) business model, which makes available to the public data collected from sensors. Consequently, companies 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^2aaS 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 business corporation, sustains costs to recruit and compensate users for their involvement. The

- *A. Capponi and P. Bouvry are with the University of Luxembourg, Luxembourg. E-mail: name.surname@uni.lu*
- *C. Fiandrino is with Imdea Networks Institute, Leganés, Madrid, Spain. E-mail: {claudio.fiandrino}@imdea.org. Claudio developed the current work as a Ph.D. student at the University of Luxembourg.*
- *D. Kliazovich is with ExaMotive, Luxembourg. E-mail: kliazovich@ieee.org.*
- *S. Giordano is with the University of Pisa, Italy. E-mail: stefano.giordano@iet.unipi.it*

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

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 these years, the research community has put lot of effort in developing incentive mechanisms to foster user participation [10], [11] and in investigating privacy issues [12], [13]. While privacy concerns are certainly a limiting factor, the cost of sensing often defines the level of user contribution and cost-effective solutions are a powerful incentive to stimulate user participation [14]. On one hand, mobile devices are battery constrained and it is important to use all available energy wisely, i.e., refrain from unnecessary sensing operations. On the other hand, reporting collected samples using wireless communication technologies, such as 3G/LTE/4G, WiFi or Bluetooth, affects battery lifetime [15], [16] and has associated data plan costs.

In this paper, we propose a novel distributed framework for data collection in opportunistic MCS systems. The proposed framework aims to minimize the cost of sensing and reporting by estimating the utility of performing such operations in a distributed manner, or the *sensing potential*. At the same time, the quality of contributed information for the system is guaranteed by using *data collection utility*. The data 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 from contributing excessively, e.g., draining the user battery completely for continuous use, or contributing 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 with a corresponding computing overhead. The synopsis of contributions of this work is as follows:

- Proposal of a novel distributed and cost-effective framework for data collection in opportunistic MCS systems. The framework maximizes quality of sensed data and minimizes sensing costs of the system and individual participants.
- Design of a custom simulator to capture crowd sensing activities in large-scale urban scenarios.
- Performance evaluation of the proposed framework analytically as well as through simulations.

The rest of the paper is organized as follows. Section 2 presents background on data collection in MCS and motivates the need for efficient data collection frameworks. Section 3 proposes a novel framework for data collection and Section 4 presents performance evaluation and results obtained from simulations. Finally, Section 5 concludes the work and outlines future research directions.

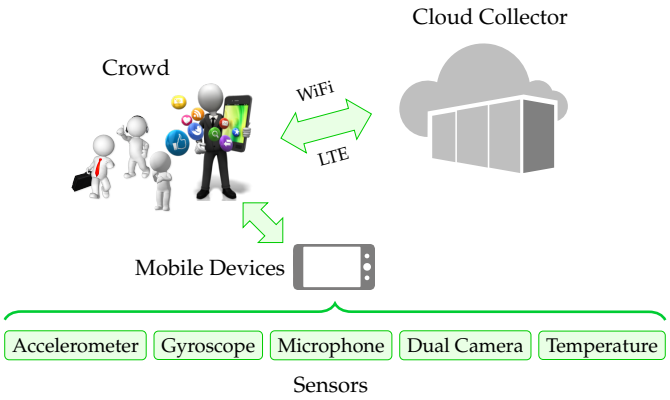


Fig. 1. Cloud-based MCS system

2 Background and Motivation

In MCS data collection architectures the participants contribute information from mobile devices' sensors.[1] This information is then delivered to a collector, typically located in the cloud, for data processing and analysis. The users deliver collected data using cellular 3G/LTE/4G or WLAN interfaces. Therefore, it becomes important to understand and assess the costs of sensing and data reporting for individual users as well as data collection capacity of the system, while maximizing the utility of the collected information. Fig. 1 illustrates the main elements of the MCS system.

A data collection framework defines the set of steps necessary to produce and deliver the information from the participants to the collector. Existing frameworks are all general-purpose or application-specific. Application-specific frameworks target only one type of application at the time. To illustrate with few examples, GasMobile [17] and NoiseMap [18] allows monitoring air and noise pollution respectively and DietSense [19] fosters healthy eating by collecting information on the type and location of consumed food. On the other hand, the salient feature of general-purpose frameworks is the capability of serving many applications at the same time. Examples of general-purpose frameworks are CARDAP [20], Medusa [21], BLISS [22], EffSense [23] and Honeybee [24].

CARDAP [20] is a general-purpose framework which enables efficient data analytics performed in a distributed fashion in fog computing platforms. The fog allows CARDAP to extend and augment functionality of a previously proposed general-purpose framework called CAROMM [25]. Many research works have focused on participatory frameworks. BLISS [22] is an online learning algorithm for data collection, where collector assigns tasks having a limited budget at disposal for rewarding users. It assumes a certain minimum number of participants who are actively engaged in sensing, while the quality of contributed data can vary. Liu et al. [26] propose an efficient user selection method for participatory type of mobile crowd sensing. The users are dynamically selected on the basis of their potential and willingness to contribute data. The potential for contribution is computed taking into account the amount of energy remaining at their mobile devices. Tasks are assigned in order to minimize

1. In the remaining part of the paper, we use terms crowd, participants and users interchangeably.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

the rejection rate, i.e., the probability that a user refuses to accomplish the assigned task. Wang et al. [23] propose an energy-efficient algorithm for uploading the sensed data. It groups users into two categories. In the first category, the users pay for the data they consume to the operators and their goal is to minimize the energy cost during data uploading. In contrast, the users belonging to the second category aim to minimize the cost of data uploading exploiting free-of-charge networks such as WiFi or Bluetooth. Similar to Piggyback CrowdSensing [14], data uploading is performed when the user places voice calls. To balance the workload among the participants while maximizing the utility of data collection, in [27] the authors propose a Nash bargaining approach where two objectives, such as load distribution and utility maximization, are considered as two cooperative players in the game. Fernando et al. [24] propose Honeybee, which is a general-purpose framework where users exploit their experience in task classification and share computing resources.

The participatory way of sensing, implemented by the aforementioned frameworks, has the disadvantage of explicitly requiring users to perform certain tasks, which may be unrealistic in many MCS systems. The opportunistic approach to sensing overcomes this limitation and avoids sending explicit requests to perform actions to the users. Hassani et al. [28] proposed Context-Aware Task Allocation (CATA), which is a framework for opportunistic MCS systems allocating tasks to users through an recruitment policy. The policy aims at selecting most appropriate users to fulfill sensing by determining similarities between the participants and the tasks. The policy considers energy consumption of sensing and data delivery operations to determine the eligibility of the participants, similarly to [29]. Moreover, this policy protects also privacy, e.g., the platform never asks the users to reveal sensible information such as user location. Han et al. [30] propose an opportunistic online distributed algorithm allowing devices to make decisions whether to perform sensing through network optimization techniques and schedule the distribution of correlated tasks among the smartphones. In addition, the algorithm maximizes utility of the sensed data while having a constrained budget cost.

Assessing the quality of collected data is important in sensor networks [31]. For MCS systems, Di Stefano et al. [32] have proposed a non-markovian stochastic Petri net formulation to evaluate the performance of MCS systems assessing Quality of Service (QoS) metrics. However, the quality of contributed information itself is not taken into account.

In this paper we propose a general-purpose framework for data collection in opportunistic cloud-based MCS systems. Unlike other proposals that focus on participatory systems [22], [23], [26], the proposed solution is for opportunistic architectures where the users do not receive specific tasks to accomplish. Instead, the mobile devices decide to perform sensing and reporting on the basis of the energy cost, the environmental context and a feedback from the cloud collector to guarantee quality of contributed information. For example, if for smartphones s_1 and s_2 the cost of sensing and reporting is high, s_1 can decide to remain idle and save energy if its samples do not bring utility for the collector, while s_2 can contribute if its samples bring high utility for the collector. Consequently, similarly to [30], these decisions are taken in a distributed manner, avoiding the need for a central intelligence that selects participants and assigns sensing tasks to them.

TABLE 1
Symbols list and description

Symbol	Description
s	Sensor s
$\mathcal{S}$	Set of sensors
a	Monitored area
$\mathcal{A}$	Set of monitored areas
t	Timeslot
d_s^a	Data collection utility for sensor s in area a
sp_s	Smartphone sensing potential for sensor s
C_s	Environmental context for sensor s
$N_s^a\|_t$	Number of samples generated by sensor s in timeslot t in area a
E_s	Energy spent by sensor s
E_s^c	Energy spent by sensor s for sensing
U_s	Utilization context for sensor s
E_s^r	Energy spent by sensor s for reporting
P^W	Power consumption of WiFi
P^L	Power consumption of LTE
B	Remaining battery charge
D	Amount of the reported data
D_s	Amount of the reported data from sensor s
D_s^W	Amount of the reported data via WiFi
D_s^L	Amount of the reported data via LTE
δ	Threshold for making sensing and reporting decisions
δ_b	Contribution of the level of battery to δ
δ_d	Contribution of the amount of reported data to δ

3 Data Collection Framework

Data collection in cloud-based opportunistic MCS systems must be cost-effective for participants and minimize the energy they spend while sensing and reporting. In addition to the efficiency, the proposed data collection framework ensures system performance by maximizing utility of the collected samples. To achieve this objective, the data collector periodically broadcasts information about the most urgently needed samples, defined as the data collection utility. Then all participants can determine whether they can contribute the requested data taking into account their cost of sensing and reporting and on the basis of the environmental context [33], [34]. For example, capturing a picture while the smartphone is in a pocket does not bring any utility for the MCS system. Table 1 lists description of symbols used in the following sections.

3.1 Data Collection Policies

The mobile devices decide to perform sensing and reporting independently one from another one. 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 δ which defines the amount of contribution each participant provides. 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 δ is illustrated in more details respectively in Subsection 3.2, Subsection 3.3, Subsection 3.4, and Subsection 3.5.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

The cloud collector and system participants may have contrary objectives. For example, the collector may require samples from the temperature sensor in a given area, while the participants of that area may be reluctant to contribute data with the goal of preserving their resources. To unify and resolve contradictions, the framework specifies two different data collection policies, a *collector-friendly* policy and a *smartphone-friendly* policy. The former puts more emphasis and prioritizes the *data collection utility*. The latter policy prevents sensing operations when the environmental context is unfavorable with an objective of saving battery power.

Collector-friendly Policy (CFP): The policy is formally defined as follows:

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

where γ is a balancing coefficient which assumes real values in the range $[0, 1]$. The parameter γ is computed by the collector and broadcasted to the participants. High values of γ give more relevance to the *smartphone sensing potential*, while low values of γ make the *data collection utility* term dominant. When the context is unfavorable ($C_s = 0$), the smartphones can still sense and report data with the CFP policy. Indeed, they receive feedback on d_s^a from the collector and compute $((1-\gamma) \cdot d_s^a)$, which is the contribution of the data collection utility to the scoring function. When the contribution is above δ, the smartphone performs both sensing and reporting.

Smartphone-friendly Policy (SFP): The policy is defined as follows:

$$C_s \cdot [\gamma \cdot sp_s + (1-\gamma) \cdot d_s^a] > \delta. \tag{2}$$

With respect to the *collector-friendly* policy defined in (1), when the environmental context is unfavorable, the devices never perform sensing and reporting, assuming the utility of such samples would be low for the collector. Unlike (1), when $C_s = 0$ the contribution given by the first member of (2) is never above the threshold δ. As a consequence, SFP prevents the devices of performing useless operations and allows to conserve resources. On the other hand, adoption of the *collector-friendly* policy would make the devices to perform sensing and reporting even if the environmental context is unfavorable. As a result, the collector receives significantly more data. The CFP can be employed to monitor phenomena that require many samples to be captured for long time periods or to boost contribution from areas that did not contributed yet enough information.

3.2 Data Collection Utility

Following the Sensing-as-a-Service (S^2aaS) paradigm, which recently gained significant attention from the research community [35], [36], [37], the data collector is located in the cloud (see Fig. 1). 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. Then, the information is delivered to the cloud, stored and made available to the cloud users. Such model perfectly suits MCS and it is of special interest when applied to smart cities [38].

Based on sensing interest, which is assumed to be known in advance, the 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. All these parameters serve as basis to identify the utility function for data collection. Because of the high number of participants contributing data, Han et al [30] define the utility of new samples to be inversely proportional of the number of already collected samples. According to this definition, also called *marginal effect*, during a sensing campaign the last sample collected does not bring any utility to the system. However, the marginal effect in data collection applies only in case a given phenomenon has to be captured only once. When performing continuous monitoring, a higher number of samples in cloud collector gathered from a long time period leads to better system accuracy [39]. Areas with higher user participation will satisfy the demand for samples faster leading to low utility in gathering more information. On the other hand, if only few samples are available from a given area the demand for additional samples is high.

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 the sensor s in the area a during the timeslot t through the Exponential Weighted Moving Average filter (EWMA):

$$\overline{N}_s^a|_t = \sigma \cdot N_s^a|_t + (1-\sigma) \cdot \overline{N}_{s-1}^a|_t, \tag{3}$$

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 σ is the exponential weighting coefficient. High values of σ 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:

$$d_s^a = \frac{1}{1 + \mathrm{e}^{-\frac{\varphi_s}{\rho_s} \cdot (-\overline{N}_s^a|_t + (1 - \frac{\rho_s}{2}))}}, \tag{4}$$

where φ_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 a using periodically by transmitting beacon messages.

3.3 Smartphone Sensing Potential

From the collector point of view, the smartphone sensing potential defines the best candidate among participants to perform sensing and reporting. Suppose two smartphones s_1 and s_2 are located in the same area and receive from the collector the same indication on the data to be collected and delivered. Then, if the cost in performing such operations is lower for s_1 than s_2, the framework should favorite contribution from s_1. The energy cost E_s each smartphone

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

s spends for both sensing and reporting describes the smartphone sensing potential.

The smartphone sensing potential sp_s is computed by each device individually and is a function of locally spent energy E_s. Similarly to the data collection utility, the relation between sp_s and E_s is described with a sigmoid function:

$$sp_s = \frac{1}{1 + \mathrm{e}^{-\frac{\zeta_s}{\theta_s} \cdot (-E_s + (1 - \frac{\theta_s}{2}))}}. \quad (5)$$

The smartphone sensing potential sp_s can assume real values in the range $[0, 1]$ and the parameters $1 - \theta_s/2$ and ζ_s control position of the center and the speed of the incline respectively. On one hand, a steep increase of the incline reduces the 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 to contribute data can be attributed to sensing (E_s^c) and reporting (E_s^r) operations:

$$E_s = E_s^c + E_s^r. \quad (6)$$

For sensing, the contribution E_s^c has to be taken into account only if the sensor s is not already in use by another application, which is defined by the *utilization context* of s, U_s. Otherwise, similar to [14], the energy spent by s is equal to zero. E_s^c is defined as follows:

$$E_s^c = \overline{E}_s^c \cdot U_s, \quad (7)$$

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

$$U_s = \begin{cases} 0, & \text{if the sensor } s \text{ is used by another application;} \\ 1, & \text{otherwise.} \end{cases} \quad (8)$$

Data reporting consists of delivering the information collected from the set of sensors $\mathcal{S}$ and transmitted 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:

$$E_s^r = \begin{cases} E^W, & \text{if WiFi or both WiFi and LTE are enabled;} \\ E^{L_1}, & \text{if WiFi is disabled and LTE is idle state;} \\ E^{L_2}, & \text{if WiFi is disabled and LTE is connected state.} \end{cases} \quad (9)$$

Most of the operating systems for smartphones, including Android and iOS, tend to prefer WiFi over cellular connectivity for data transmission, as it is more energy efficient [40] and users do not consume the data plan they pay to the cellular operators [41]. As a result, when both WiFi and LTE interfaces are active, transmissions take place via WiFi. The energy E^W spent during the transmission time τ_{tx} is defined as:

$$E^W = \int_0^{\tau_{tx}} P_{tx}^W \, \mathrm{d}t, \quad (10)$$

where P_{tx}^W is the power consumed for transmissions of WiFi packets generated at rate λ_g [42]:

$$P_{tx}^W = \rho_{id} + \rho_{tx} \cdot \tau_{tx} + \gamma_{xg} \cdot \lambda_g. \quad (11)$$

The parameters ρ_{id}, ρ_{tx} and γ_{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 [43] 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* [40], 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 coming back to *idle*. In this case, the energy consumption for the smartphone during reporting can be defined as:

$$E^{L_1} = P_{\mathrm{P}} \cdot T_{\mathrm{P}} + P_{\mathrm{tx}}^L \cdot T_{\mathrm{tx}} + P_{\mathrm{tx}}^L \cdot DRX_{\mathrm{IT}} + P_{\mathrm{DRX}} \cdot RRC_{\mathrm{IT}}, \quad (12)$$

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:

$$E^{L_2} = \int_0^{T_{\mathrm{tx}}} P_{\mathrm{tx}}^L \, \mathrm{d}t. \quad (13)$$

The power consumption for transmitting data P_{tx} is given by the model of [40]:

$$P_{\mathrm{tx}}^L = \alpha_{\mathrm{ul}} \cdot T_{\mathrm{ul}} + \beta, \quad (14)$$

where T_{ul} represents the uplink throughput and the parameters α_{ul} and β are the power spent during transmission and the base power respectively [40].

When the value of the last sensed sample l_s^x is close to the last reported value l_s^r, avoiding reporting allows to save energy for the devices and improves the utility of data collection at the collector. More in details, data reporting does not take place if:

$$|l_s^x - l_s^r| < \epsilon_s, \quad (15)$$

where ϵ_s is a threshold defining the similarity between the two values and is application-dependent. For real implementation, proper setting of the parameter ϵ_s requires a control layer to manage and synchronize the sensing activity of the mobile device.

This operation can provide significant energy savings, especially for sensing data that do not exhibit high variability. This is typical for meteorological measurements, such as temperature, atmospheric pressure and humidity. When the variation between the values of samples generated from the same sensor s in subsequent timeslots is minimal, the utility of reporting the last sample is low. Therefore, to maximize system-level utility and save energy, the framework prevents

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

reporting and stops sensing. To model this procedure, the next timeslot is defined as:

$$t_{\text{next}} = \begin{cases} t + i \cdot n, & \text{if } |l_s^x - l_s^r| < \epsilon_s; \\ t + 1, & \text{otherwise.} \end{cases} \quad (16)$$

The parameter n is a fixed number of timeslots to skip. The initial value of n is set to 1 and then doubled every time up to the maximum of 1024 [44]. The parameter i is the number of times the variation between the values of the samples remains below ϵ_s. In case the decision is to not report, the cost of reporting E_s^r assumes values equal to zero. It is worth to highlight that (16) determines t_{next} for each sensor s individually, i.e., the correlation between sensors, which is application-dependent, is not captured by the model.

3.4 Profiling the Environmental Context

Environmental context C_s of sensor s defines the set of facts and circumstances happening around the mobile device. It accounts for both the location of a mobile device (e.g., lying on a table, in a pocket, in a hand, indoor or outdoor) and its mobility pattern (e.g., stationary or on the move, with the possibility to recognize the type of movement). The environmental context is usually detected with always-on sensors, such as the accelerometer, enabled with recognition of movement patterns (e.g., walking or running [45]) and actions (e.g., driving, riding a car or sitting [46]). 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. Although a few practical solutions exist [47], estimating the environmental context is not simple.

The utility of each sample is defined according to the environmental context. For the sake of simplicity and without loss of generality, we assume C_s to take binary values:

$$C_s = \begin{cases} 1, & \text{if the sample contributes to the sensing objective;} \\ 0, & \text{otherwise.} \end{cases} \quad (17)$$

When starting a new sensing campaign, the collector first broadcasts profiles defining the environmental context to all the devices. Each sensing campaign is tailored to serve a different application, for instance air monitoring or gluten detection. Therefore, the environmental context profiles have to be different and tailored to the application in use.

3.5 Threshold for Sensing and Reporting

Sensing and reporting operations occur when *data collection utility* and *smartphone sensing potential* are greater than a threshold δ, which means that the mobile devices sustain a cost to produce useful data for the cloud collector. In addition, comparing *data collection utility* and *smartphone sensing potential* with δ prevents the users from contributing too much or too little. The threshold δ is computed locally at the mobile device. Proper setting of this parameter is essential to define the amount of data each device opportunistically contributes. To define δ, we take into account the actual *level of the battery* of the devices (denoted as B) and the *amount of reported data* the devices have already contributed to the system (denoted as D). These two parameters define historical cost experienced by the devices. The parameter δ is defined as follows:

$$\delta = f(\delta_b, \delta_d); \quad (18)$$

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

$$f(\delta_b, \delta_d) = \frac{\delta_b + \delta_d}{2}. \quad (19)$$

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 suggests that the relation between B and δ_b follows a negative exponential law. As a result, low values of B will make δ_b to assume values close to 1. $B_{\min}$ is the minimum level of the battery under which the device stops contributing. Note that $0 \leq B_{\min} < B$. As a consequence, δ_b is defined as follows:

$$\delta_b = \alpha^{\lambda \cdot B}, \quad B_{\min} \leqslant B < 1. \quad (20)$$

Fig. 2(a) shows the function. The parameter α can assume arbitrary real values between $[0, 1]$, and $\lambda > 1$.

Users can tune the parameter $B_{\min}$. Proper setting of this parameter plays a crucial role. Values of $B_{\min}$ close to the current battery level B lead the device to contribute little amount of data. On the other hand, setting $B_{\min}$ so that the difference $B - B_{\min}$ is close to 1, will make the device to contribute for longer periods of time.

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 LTE (D_s^L) connections and is contributed by the set of sensors $\mathcal{S}$ of a single device. D is defined as follows:

$$D = \sum_{s \in \mathcal{S}} D_s^W + D_s^L. \quad (21)$$

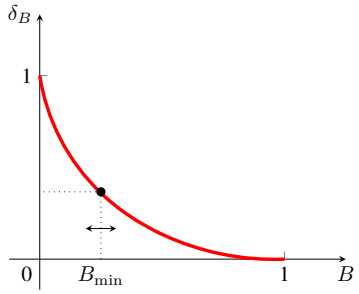
The more data the device contributes, the higher the values parameter δ_d assumes. As a result, the devices that have already delivered a significant amount of data to the cloud collector in the past contribute further only if it is necessary. On the other hand, low values of the parameter δ_d will enforce the contribution of the devices that have provided little contribution to the system. The relation between D and δ_d is modelled as follows:

$$\delta_d = \begin{cases} 1, & \text{if } D \geq D_{\max}; \\ \log\left(1 + \dfrac{D}{D_{\max}}\right), & \text{otherwise;} \end{cases} \quad (22)$$

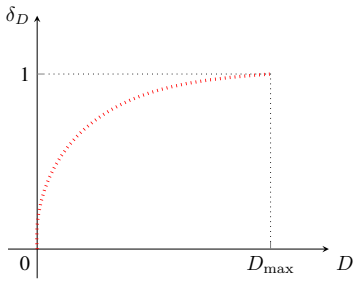
where $D_{\max}$ is the maximum amount of data each device is willing to deliver (see Fig. 2(b)). This parameter can be tuned by the users periodically.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing



(a) Contribution of the remaining charge of the battery



(b) Contribution of the amount of reported data

Fig. 2. Parameters affecting the threshold for sensing and reporting δ

TABLE 2
Sensor and communication equipment parameters used for performance evaluation

(a) Sensor Equipment

SENSOR	PARAMETER	VALUE	UNIT
Accelerometer	Sample rate	50	Hz
	Sample size	12	Bits
	Current	35	µA
Temperature	Sample rate	182	Hz
	Sample size	16	Bits
	Current	182	µA
Pressure	Sample rate	157	Hz
	Sample size	16	Bits
	Current	423.9	µA

(b) Communication Equipment

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

4 PERFORMANCE EVALUATION

This section illustrates performance evaluation of the proposed framework for data collection. We provide an analytical study and investigate the effectiveness of the framework in a large scale city-size scenario with thousands of participants.

4.1 Analytical Results

4.1.1 Set-up

For the evaluation, four users contribute data and each equipped with one device only. The devices have different level of remaining battery charge: 10%, 25%, 50% and 100%. The battery charge is one of the parameters affecting the threshold δ for sensing and reporting (parameter δ_b). We compute δ_b using (20) by setting $\alpha = 0.7$ and $\lambda = 10$ after having performed analytical analysis that for space reasons we omit. The analysis aims to quantify the amount of contribution at high and low levels of battery charge. The other parameter, δ_d, corresponds to the amount of previously reported data and is set to 0 for all the users at the beginning of evaluation. The evaluation period is 600 timeslots and each of them lasts 1 second, which corresponds to a 10 minutes long evaluation period.

Evaluated mobile devices are equipped with a number of sensors including accelerometer, temperature and pressure sensors, and transmit information using WiFi. Without loss of generality, having only one communication technology simplifies understanding of properties of the data collection framework. For the sensing equipment, the devices exploit real sensors implemented in current smartphones and tablets. Specifically, the FXOS8700CQ 3axis linear accelerometer from Freescale Semiconductor [48] and the BMP280 from Bosch [49], which is a digital pressure and temperature sensor. Equation (11) describes WiFi power consumption spent by the devices for communication. Table 2 presents the detailed information on communication and the parameters.

For the sake of simplicity, all the users obtain the same profiles for the environmental context from the collector, e.g., $C = \{1.0, 1.0, 1.0\}$. During each timeslot t, the devices generate the current context $C_t = \{x, y, z\}$, where x, y and z define the context for the accelerometer, temperature and pressure sensors. The parameters are generated as random variables following exponential distribution with rate 1.25. Then, C_t is compared with C. Values of C_t close to 1.0 indicate that the smartphone performs sensing in favorable conditions. On the other hand, values of C_t close to 0.0 correspond to samples generated in unfavorable conditions. Therefore, such data does not bring any utility from the collector point of view. The utilization context U_s is modelled as a random variable that assumes values uniformly distributed in the range $[0, 1]$ and is generated during each timeslot for each user device.

4.1.2 Results

The main performance characteristics of the proposed framework include the amount of collected data and the energy cost for sensing and reporting. In addition, we evaluate the impact of the main framework control parameters: the threshold for sensing and reporting δ and the coefficient γ balancing the impact of the data collection utility and the smartphone sensing potential. The latter parameters are calculated per timeslot.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

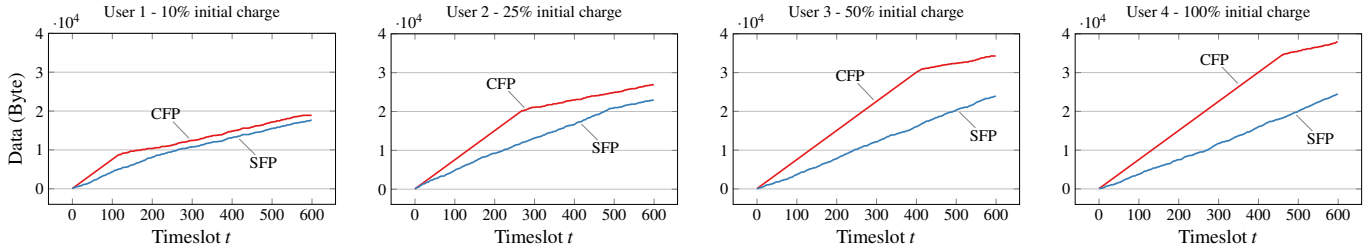
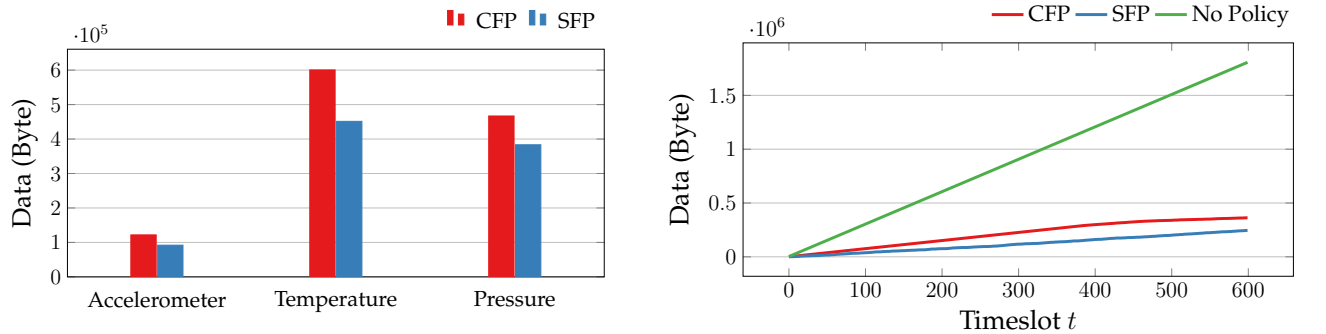


Fig. 3. Amount of data per user collected from the accelerometer



(a) Analysis per sensor

(b) Comparison of policies

Fig. 4. Total amount of collected data

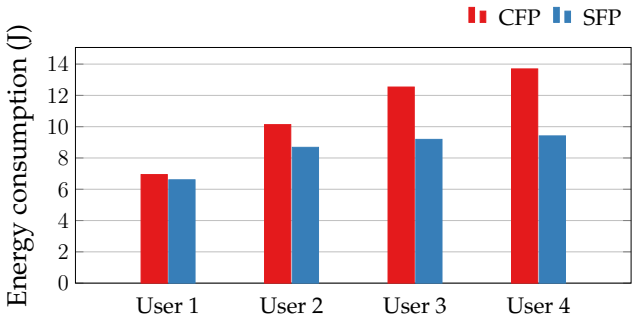


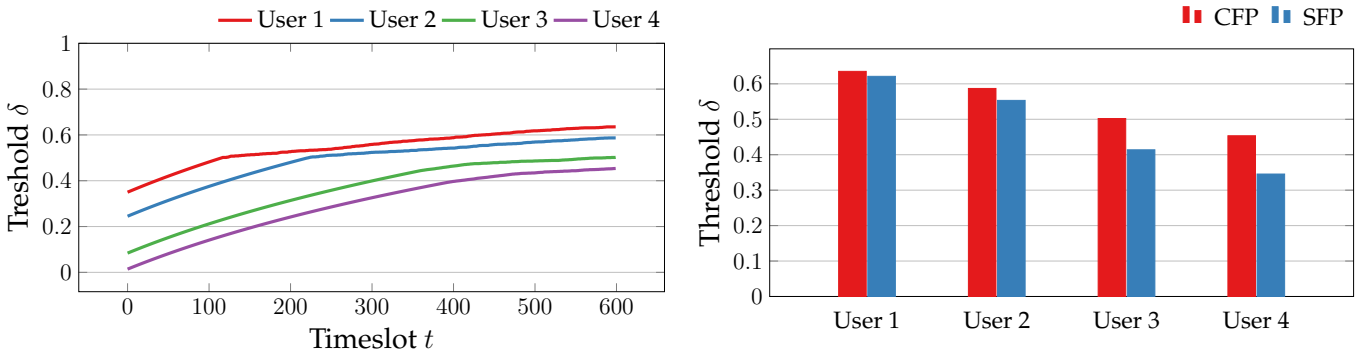
Fig. 5. Energy consumption for the data collection policies

Amount of Collected Data: For the analysis, the parameter γ is fixed to $1/2$ to give equal importance to the data collection utility and the smartphone sensing potential. Fig. 3 illustrates the amount of data generated by four users from accelerometer readings. Other sensors exhibit a similar behavior and, for space reasons, the results are not shown. The performance of both collector- and smartphone-friendly policies are investigated for the entire evaluation period. The CFP policy fosters users to contribute larger amounts of data than the SFP policy. This behavior is expected, as the SFP is tailored to perform operations only when the environmental context guarantees the devices to generate useful information to the system. Interestingly, User 1 contributes nearly the same amount of data with both policies. Having the lowest remaining battery charge, the threshold δ assumes high values and limits user contribution. With $\gamma = 0.5$, the contribution of data collection utility and the smartphone sensing potential are equal for taking sensing and reporting decisions. As a result, User 1 reports measurements only when the collector indicated having high interest in obtaining samples from this user. For all the other users, the amount of contributed data has a sharp increase, which becomes smoother towards the end of the evaluation period. The change in speed of the curves depends on the initial remaining charge of the battery. The motivation is twofold. On one hand, having obtained enough samples the utility at the collector becomes smaller (data collection utility factor). On the other hand, the remaining capacities of the devices decrease with time while the amount of already reported data increase. As a result, it becomes more difficult to meet the threshold δ and this lowers the contribution rate.

Fig. 4 analyzes the total amount of collected data. In particular, Fig. 4(a) compares the amount of information each sensor generates. The accelerometer generates the lowest amount of data in comparison with the temperature and pressure sensors. Obviously, the amount of generated data is proportional to the sampling rate and resolution size, with the temperature sensor having the highest rate and resolution size (see Table 2(a)). Low values of sampling frequency make the difference between the total amount of generated data by the two policies small. On the other hand, with the increase of the sampling frequency such difference increases. Fig. 4(b) provides a comparison between the total amount of data generated by the two policies and when no policies are utilized. Each curve denotes the cumulative amount of data generated by the three sensors together. As expected, the total amount of data generated without using the proposed framework is higher than having active either CFP or SFP. This is because the devices perform sensing and reporting even when the environmental context C_t is unfavorable. As a result, many of the generated samples do not bring any additional utility to the system. The SFP makes the user to only perform operations when the environmental context is favorable and for this reason generates the lowest amount of data. On the other hand, the CFP is more flexible and allows the devices to sense and report data even in case the environmental context is unfavorable, but the data collection

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing



(a) Threshold δ during the period of analysis

(b) Comparison between data collection policies

Fig. 6. Analysis of threshold δ

utility and the smartphone sensing potential conditions should always be met. As a consequence, the amount of data generated by means of the CFP is significantly lower than the case when the framework is not utilized.

Energy Cost: Fig. 5 compares energy consumption of each user for CFP and SFP policies. The results are obtained at the end of the evaluation period of 600 timeslots. As expected, User 1 consumes the least amount of energy. Indeed, s/he has only 10% of remaining battery charge at the beginning of the experiment and the framework tries to limit its contribution. For User 1, both CFP and SFP provide a nearly equal energy consumption. Thus, it is possible to conclude that under the current setting, the framework makes users with low remaining charge to contribute only when the environmental context ensures that the collected data brings utility for the system. For all the other users, the difference between the energy spent under the two policies is increasing with the increase of the remaining battery charge at the beginning of the evaluation period. The overall energy spent under is nearly 14 J and for a smartphone equipped with a 2550 mAh and 3.85 V battery such as the Samsung Galaxy s6, the total energy at the disposal is as of 35343 J. As a consequence, it is possible to conclude that contributing data opportunistically to a MCS system does not affect dramatically the performance of today mobile devices.

Analysis of Data Collection Control Parameters: Having evaluated the amount of data contributed and the cost to generate such information, we now study the impact of the parameters δ and γ on the overall performance. Fig. 6 analyzes the behavior of the threshold δ that enables the device to decide whether to perform sensing and reporting. For this analysis, the parameter γ is set to 0.5. Fig. 6(a) shows the values δ assumes during the evaluation period for all the users under the CFP policy. For all the users, δ increases with time. While contributing data, on one hand the users spend energy and this affects the remaining charge of the battery (parameter δ_b). On the other hand, delivering data increases the amount of already reported data (parameter δ_d). As expected, for User 1 the threshold δ is always higher than for all the other users because her remaining battery charge at the beginning of the experiments is the lowest. However, unlike the other users, the speed of increase of the curve of User 1 is slow. Indeed, all the other users contribute more data than User 1. As a result, the contribution of the amount of already reported data for User 1 is lower than for User 2, 3 and 4. Fig. 6(b) compares the values δ assumes at the end of the evaluation period for both policies CFP and SFP. Unlike the results obtained for energy consumption (see Fig. 5), for all the users the difference between the values δ assumes with the two policies decreases. Similarly to Fig. 5, for User 1 such difference is small. Intuitively, higher values of δ limit user contribution and correspond in achieving low energy consumption.

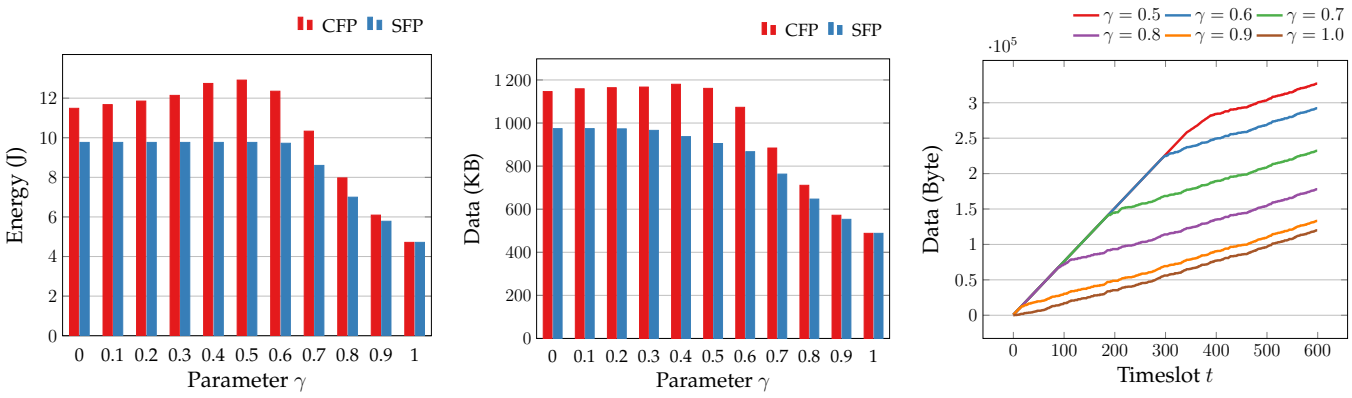
Fig. 7 studies the impact of the parameter γ on the performance of the framework. Recalling (1) and (2), the parameter γ defines the weight of the data collection utility and the smartphone sensing potential for taking decisions to perform sensing and reporting. Fig. 7(a) and Fig. 7(b) show respectively the amount of data generated and the energy the devices consume for both CFP and SFP policies with increasing values of γ. Low values of γ give more importance to the data collection utility, hence the devices contribute data even if they experience a high energy cost as Fig. 7(a) highlights. On the other side, with the increase of the parameter γ, the framework is tailored to pay more attention to the potential the devices have in perform sensing. This translates in a better energy management at the cost of delivering a lower amount of data. For the CFP policy, notice that the energy cost is increasing smoothly for values of γ below 0.5 (see Fig. 7(a)) although the amount of data generated remains almost constant (see Fig. 7(b)). This energy cost should be attributed to the utilization context U_s. When the parameter γ assumes values higher than 0.6 both the energy and the amount of collected data have a sharp decrease. As the relevance of the data collection utility decreases, the smartphone sensing potential factor becomes crucial in order to take sensing and reporting decisions. To better analyze such decrease, we plot in Fig. 7(c) the amount of data User 4 generates for values of $\gamma > 0.6$. Similarly to Fig. 3, in all the curves the increase is at first sharp and then becomes smooth. The higher the values γ assumes, the sooner the change of slope and the lower the amount of data the user contributes. In turns, this leads to energy savings.

4.2 Simulation Results

To evaluate and assess the efficiency of the proposed framework with a large number of participants and in a citywide

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing



(a) Energy Cost (b) Comparison amount of data collected (c) Amount of data collected by User 4

Fig. 7. Analysis of parameter γ

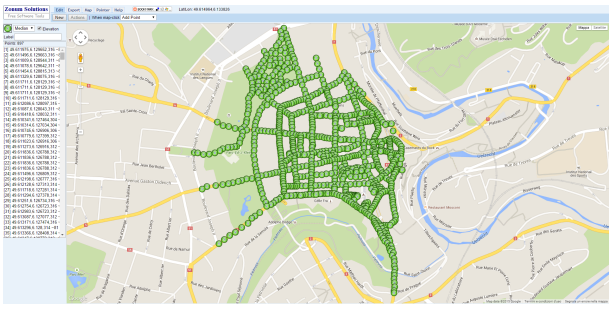


Fig. 8. Map of city center of Luxembourg from DigiPoint

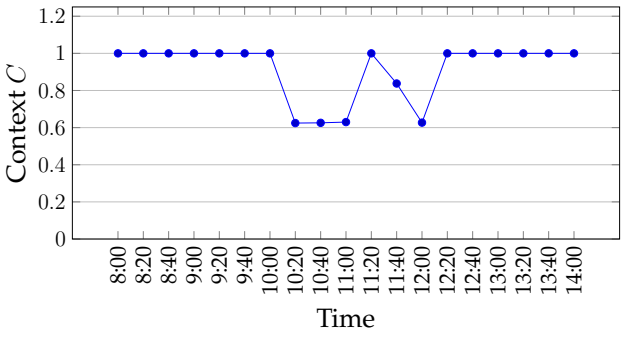


Fig. 10. Distribution of context during evaluation period

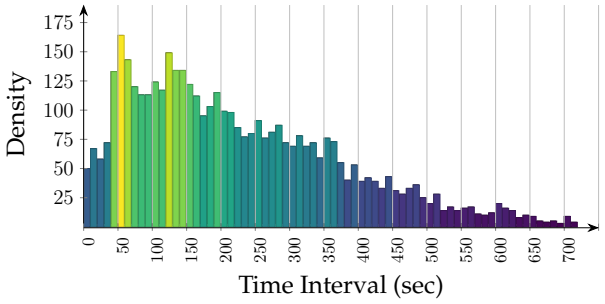


Fig. 9. User distribution of mobility trace "kth/walkers"

scenario, we have built a custom crowd sensing simulator. The following sections presents the detailed functionality along with the simulation scenario and results.

4.2.1 Simulation set-up

The developed simulator is a discrete-event simulator where the participants of the MCS system contribute data to the collector opportunistically. The simulator supports pedestrian mobility, while data generation exploits sensors commonly available in mobile devices (see Section 4.1). The simulation results can be obtained the level of individual devices as well as the system level, which helps to analyze data reporting progress and efficiency of the employed crowd sensing techniques.

The center of Luxembourg city was selected for the simulations. It covers an area of 1.11 km^2 and is the home of many national and international institutional buildings. To obtain information about the streets of the city, the simulator exploits a crowdsourced application providing free access to street-level maps[2]. Such information is given in form of a set of coordinates $\mathcal{C}$ containing $<$*latitude*, *longitude*, *altitude*$>$ (see Fig. 8).

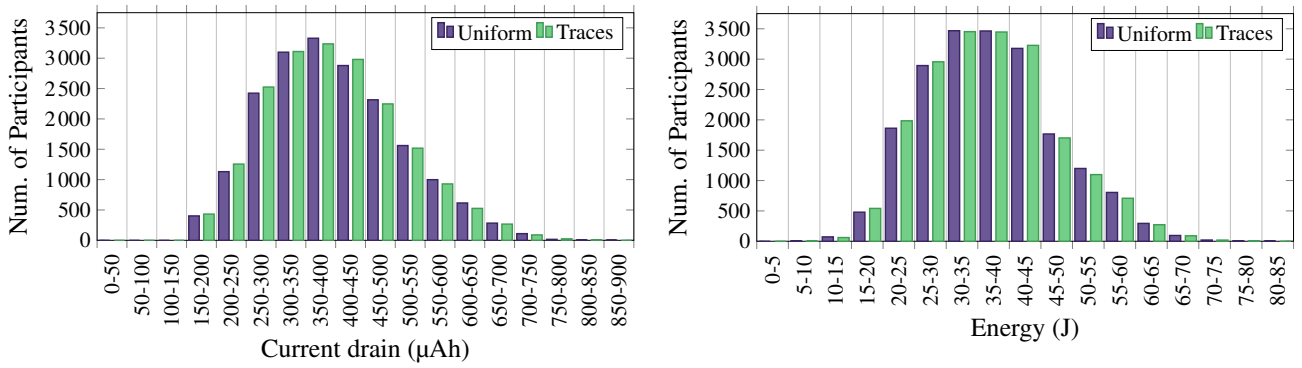
The participants move along the streets of the city, with their original locations randomly assigned form the set of coordinates $\mathcal{C}$. The number of participants is set to 20 000, which corresponds to more than one fifth of the population of Luxembourg (110,499 inhabitants as of end 2015). Each 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. The participants contribute data to the collector while walking. Once the period of walking ends, they stop moving and contributing. As a consequence, users generate information for a little period of time along the day, which allows to study the system performance under a worst case scenario. The users start walking in the city according to two arrival patterns. In the first one, for simplicity, the start time of the walk is uniformly distributed between 8:00 AM and 1:40 PM. The second arrival pattern is based on a study on a data set containing traces of pedestrian mobility (ostermalm_dense_run2) [50]. Fig. 9 shows the probability density function of the users arrival, which was adapted to our arrival time period 8:00 AM - 1:40 PM.

Fig. 10 illustrates the values the context C assumes along the simulation period. Each mark represents the average value of C during a 20 minutes interval. For instance,

2. DigiPoint: http://www.zonums.com/gmaps/digipoint.php

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing



(a) Sensing Cost

(b) Communication Cost

Fig. 11. Energy spent for sensing and communication

the mark at 10:20 AM corresponds to the average value C assumes during 10:00 AM to 10:20 AM. To study the impact of the context on the simulations, unfavorable context conditions are set in specific time periods, namely 10:00 AM - 11:00 AM and 11:30 AM - 12:00 PM.

Data generation takes place with the same set of sensors used for the analytical study. Communications occur over the WiFi link, having obtained the precise location of WiFi hotspots in form of <*latitude, longitude*> [3]. Table 2 presents the detailed information of the simulation parameters.

4.2.2 Results

For evaluation purposes, we conduct experiments aiming to assess the amount of information gathered from the devices and the energy cost consumed during data reporting.

Cost of sensing and reporting: To analyze the energy cost under a worst case scenario, the devices generate data continuously during the movement of the users.

Fig. 11 presents the distribution of users and their energy spent for sensing with the uniform and traces-based user arrival patterns. As expected, the user arrival pattern does not influence the energy consumption, which only depends on the amount of time the users generate data. As the users contribute data for time periods as low as 10 minutes up to time periods of a maximum of 20 minutes, the profiles of Fig. 11(b) and Fig. 11(a) follows a normal distribution. Current drain of sensing operations is on average 373.41 µAh and 368.80 µAh for uniform and traces-based arrival patterns. 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. 11(b)).

Amount of data collected: The amount of information reported by users devices is unveiled in the following two experiments. First, we evaluate the amount of data generated per single sensor for the two different user arrival patterns. To capture distribution of the samples per area, we divide Luxembourg City map into five areas and measure the number of samples reported from each area. The objective is to understand the amount and the variability of the data generation process along the time period 8:00 AM - 2:00 PM. Having a higher number of participants, it allows to test the performance of the framework in a large scale environment.

Fig. 12 shows the total amount of data collected along with the simulation period for the two user arrival patterns. As expected, the amount of data is proportional to the sampling frequencies of the three considered sensors. Recalling that each user contributes only during a short period of time (10 to 20 minutes), the amount of collected information is considerable. For example, 20 000 users arriving according to the uniform arrival pattern would generate 2.61797 GB, 12.71 GB and 10.96 GB for the accelerometer, temperature and pressure sensors respectively. Fig. 12(a) shows the results for the uniformly distributed arrival pattern. The impact of the unfavorable context during the periods 10:00 AM - 11:00 AM and 11:30 AM - 12:00 PM is clear and reduces the amount of generated data. Fig. 12(b) illustrates the results for the user arrival pattern based on the data set. The unfavorable context impacts the amount of generated data similarly to the previous case.

Having the knowledge on the amount of data the users can contribute is important for the applications and to determine the accuracy in mapping a phenomena. However, for drawing more precise conclusions, it is fundamental to determine also where and when these samples are generated. For this, we define a new metric, called Sample Distribution (SD), which measures the amount of generated samples per meter and is defined as follows:

$$\text{SD} = \frac{N_s^a|_t}{\overline{\Delta}}, \tag{23}$$

where $\overline{\Delta}$ is the average distance between samples and $N_s^a|_t$ is the number of samples generated (see Table 1). The parameter $\overline{\Delta}$ is defined as follows:

$$\overline{\Delta} = \frac{\sum_{\substack{i,j \\ i \geq j}}^{n} d(i,j)}{\frac{n(n-1)}{2}}. \tag{24}$$

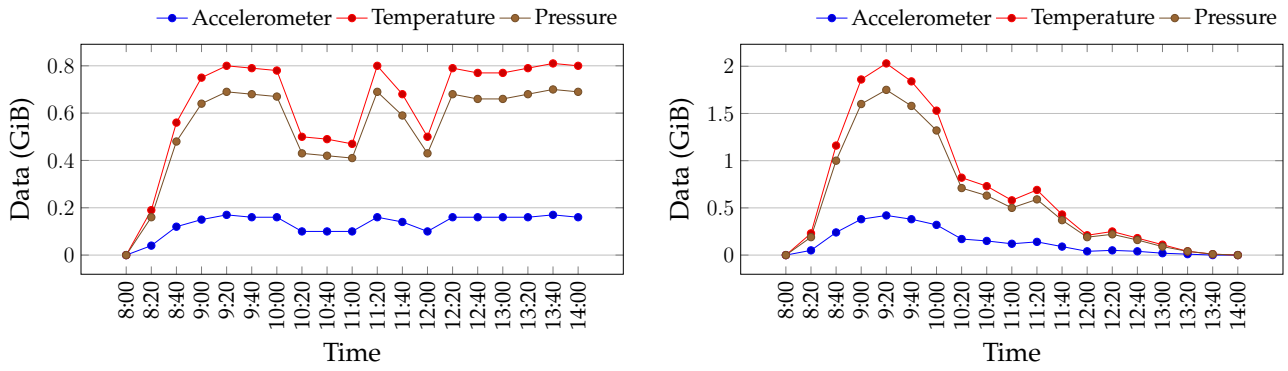
The term $d(i,j)$ is the distance (in meters) between the location where the samples i and j were generated.

Fig. 13 plots SD for each segment of the city map for the time period 10:00 AM - 10:30 AM. In this analysis, only accelerometer samples were utilized. The SD metric weakly

3. Online: https://www.hotcity.lu/en/laptop/www/About/Wi-Fi-coverage

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing



(a) Arrival pattern with uniform distribution

(b) Arrival pattern based on traces

Fig. 12. Amount of data generated

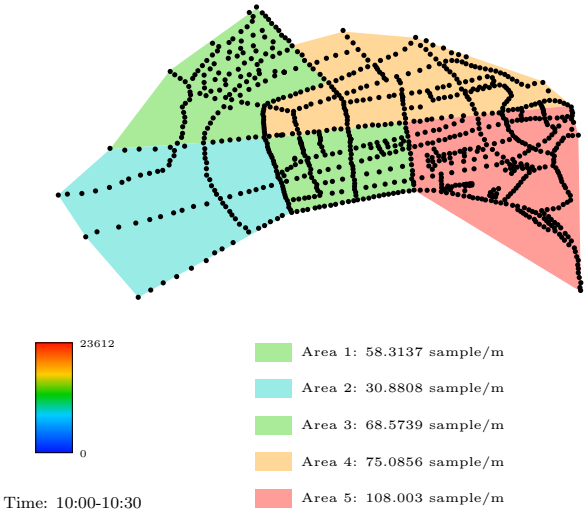


Fig. 13. Sample distribution (SD) for 20 000 users moving in the center of Luxembourg

depends on the size of the area. Although being wider than Area 3, a large part of Area 2 is a public park with a fewer number of streets. The reason is that the SD metric measures the distribution of the samples taking into account the location where they have been generated, recording latitude, longitude and altitude. Consequently, high density areas, such as Area 5, exhibit high values of SD.

Fig. 14 shows the distribution of SD for all the 5 areas for the entire simulation period. 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 end 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 low 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.

5 Conclusion

In this paper we propose a new distributed data collection framework for opportunistic cloud-based MCS systems. The framework aims at minimizing the costs for the participants in performing sensing and reporting while maximizing the utility in data collection for the cloud collector. The performance of the framework was verified both analytically and through simulations in a real urban environment and with a large number of participants. We analyze the costs the participants experience and the amount of data the system allows to gather. The analytical results highlight that the major contribution to energy consumption is attributed to reporting and not sensing. The simulation results confirm the effectiveness of the proposed approach in a real urban environment for a large number of participants. For future work, we plan to implement the current model with an application to verify the performance of the framework in real world. We also consider to extend the current model to consider caching policies for buffering samples and improve efficiency of data delivery. Furthermore, we plan to study the impact of the size of the areas to the data collection.

Acknowledgments

The authors would like to acknowledge the funding from National Research Fund (FNR), Luxembourg in the framework of ECO-CLOUD and iShOP projects.

References

[1] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Communications Magazine*, vol. 49, no. 11, pp. 32–39, November 2011.

[2] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, March 2014, pp. 593–598.

[3] H. Falaki, D. Lymberopoulos, R. Mahajan, S. Kandula, and D. Estrin, "A first look at traffic on smartphones," in *10th ACM SIGCOMM Conference on Internet Measurement*, ser. IMC, 2010, pp. 281–287.

[4] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: architecture, applications, and approaches," *Wireless Communications and Mobile Computing*, vol. 13, no. 18, pp. 1587–1611, 2013.

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

[6] 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.

[7] B. Kantarci and H. T. Mouftah, "Trustworthy sensing for public safety in cloud-centric internet of things," *IEEE Internet of Things Journal*, vol. 1, no. 4, pp. 360–368, Aug 2014.

[8] G. Cardone, A. Cirri, A. Corradi, and L. Foschini, "The participact mobile crowd sensing living lab: The testbed for smart cities," *IEEE Communications Magazine*, vol. 52, October 2014.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

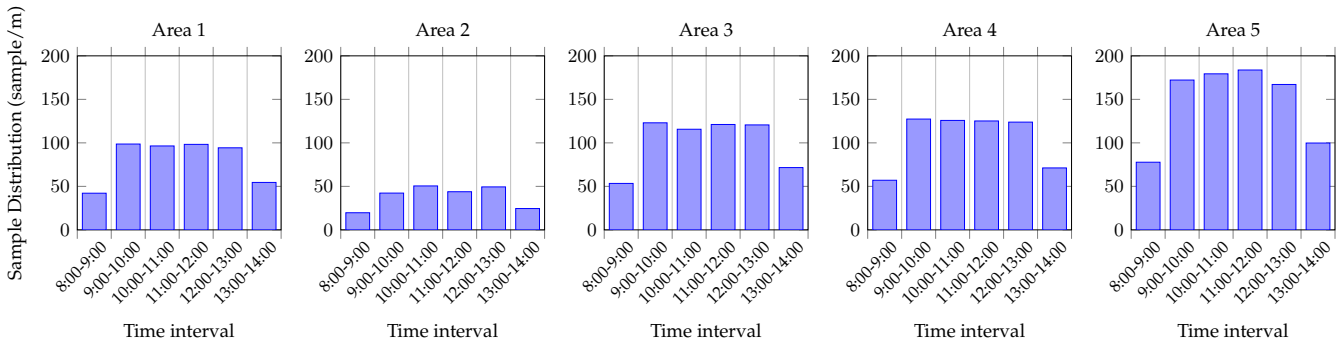


Fig. 14. Sample distribution (SD) for the different areas over the time period 8:00-14:00

[9] H. Schaffers, N. Komninos, M. Pallot, B. Trousse, M. Nilsson, and A. Oliveira, "Smart cities and the future internet: Towards cooperation frameworks for open innovation," in *The Future Internet*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6656, pp. 431–446.

[10] X. Zhang, Z. Yang, W. Sun, Y. Liu, S. Tang, K. Xing, and X. Mao, "Incentives for mobile crowd sensing: A survey," *IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 54–67, 2016.

[11] L. Jaimes, I. Vergara-Laurens, and A. Raij, "A survey of incentive techniques for mobile crowd sensing," *IEEE Internet of Things Journal*, vol. 2, no. 5, pp. 370–380, Oct 2015.

[12] D. Christin, A. Reinhardt, S. S. Kanhere, and M. Hollick, "A survey on privacy in mobile participatory sensing applications," *Journal of Systems and Software*, vol. 84, no. 11, pp. 1928–1946, 2011.

[13] L. Pournajaf, L. Xiong, D. A. Garcia-Ulloa, and V. Sunderam, "A survey on privacy in mobile crowd sensing task management," *Technical Report TR-2014-002, Department of Mathematics and Computer Science, Emory University*, 2014.

[14] N. D. Lane, Y. Chon, L. Zhou, Y. Zhang, F. Li, D. Kim, G. Ding, F. Zhao, and H. Cha, "Piggyback crowdsensing (pcs): Energy efficient crowdsourcing of mobile sensor data by exploiting smartphone app opportunities," in *11th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys, 2013, pp. 1–14.

[15] 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.

[16] R. Friedman, A. Kogan, and Y. Krivolapov, "On power and throughput tradeoffs of WiFi and Bluetooth in smartphones," *IEEE Transactions on Mobile Computing*, vol. 12, no. 7, pp. 1363–1376, July 2013.

[17] D. Hasenfratz, O. Saukh, S. Sturzenegger, and L. Thiele, "Participatory air pollution monitoring using smartphones," in *In Mobile Sensing: From Smartphones and Wearables to Big Data*. Beijing, China: ACM, Apr 2012.

[18] I. Schweizer, R. Bärtl, A. Schulz, F. Probst, and M. Mühläuser, "Noisemap - real-time participatory noise maps," in *In Second International Workshop on Sensing Applications on Mobile Phones*, 2011, pp. 1–5.

[19] 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 *4th ACM Workshop on Embedded networked sensors*, 2007, pp. 13–17.

[20] P. P. Jayaraman, J. B. Gomes, H. L. Nguyen, Z. S. Abdallah, S. Krishnaswamy, and A. Zaslavsky, "Scalable energy-efficient distributed data analytics for crowdsensing applications in mobile environments," *IEEE Transactions on Computational Social Systems*, vol. 2, no. 3, pp. 109–123, Sept 2015.

[21] M.-R. Ra, B. Liu, T. F. La Porta, and R. Govindan, "Medusa: A programming framework for crowd-sensing applications," in *10th ACM International Conference on Mobile Systems, Applications, and Services*, ser. MobiSys, 2012, pp. 337–350.

[22] K. Han, C. Zhang, and J. Luo, "BLISS: Budget limited robust crowdsensing through online learning," in *IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2014, pp. 555–563.

[23] L. Wang, D. Zhang, Z. Yan, H. Xiong, and B. Xie, "effSense: A novel mobile crowd-sensing framework for energy-efficient and cost-effective data uploading," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 45, no. 12, pp. 1549–1563, Dec 2015.

[24] N. Fernando, S. W. Loke, and W. Rahayu, *Honeybee: A Programming Framework for Mobile Crowd Computing*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 224–236.

[25] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha, "Using on-the-move mining for mobile crowdsensing," in *2012 IEEE 13th International Conference on Mobile Data Management*, July 2012, pp. 115–124.

[26] C. Liu, B. Zhang, X. Su, J. Ma, W. Wang, and K. Leung, "Energy-aware participant selection for smartphone-enabled mobile crowd sensing," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2015.

[27] J. Li, Y. Zhu, and J. Yu, "Load balance vs utility maximization in mobile crowd sensing: A distributed approach," in *Proceedings of IEEE GLOBECOM*, Austin, USA, Dec. 2014, pp. 259–264.

[28] A. Hassani, P. D. Haghighi, and P. P. Jayaraman, "Context-aware recruitment scheme for opportunistic mobile crowdsensing," in *IEEE 21st International Conference on Parallel and Distributed Systems (ICPADS)*, Dec 2015, pp. 266–273.

[29] M. H. ur Rehman, C. S. Liew, and T. Y. Wah, "Uniminer: Towards a unified framework for data mining," in *Fourth World Congress on Information and Communication Technologies (WICT)*, Dec 2014, pp. 134–139.

[30] Y. Han, Y. Zhu, and J. Yu, "Utility-maximizing data collection in crowd sensing: An optimal scheduling approach," in *12th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*, June 2015, pp. 345–353.

[31] F. Bijarbooneh, P. Flener, E. Ngai, and J. Pearson, "Optimising quality of information in data collection for mobile sensor networks," in *IEEE/ACM 21st International Symposium on Quality of Service (IWQoS)*, June 2013, pp. 1–10.

[32] S. Distefano, F. Longo, and M. Scarpa, "QoS assessment of mobile crowdsensing services," *Journal of Grid Computing*, pp. 1–22, 2015.

[33] 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.

[34] I. Carreras, D. Miorandi, A. Tamilin, E. Ssebaggala, and N. Conci, "Crowd-sensing: Why context matters," in *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, March 2013, pp. 368–371.

[35] X. Sheng, J. Tang, X. Xiao, and G. Xue, "Sensing as a service: Challenges, solutions and future directions," *IEEE Sensors Journal*, vol. 13, no. 10, pp. 3733–3741, Oct 2013.

[36] B. Kantarci and H. Mouftah, "Sensing services in cloud-centric internet of things: A survey, taxonomy and challenges," in *IEEE ICC Workshop (ICCW)*, June 2015, pp. 1865–1870.

[37] S. Chatterjee, S. Misra, and S. Khan, "Optimal data center scheduling for quality of service management in sensor-cloud," *IEEE Transactions on Cloud Computing*, vol. PP, no. 99, pp. 1–1, 2015.

[38] 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.

[39] D. Mendez, M. Labrador, and K. Ramachandran, "Data interpolation for participatory sensing systems," *Pervasive and Mobile*

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSUSC.2017.2666043, IEEE Transactions on Sustainable Computing

Computing, vol. 9, no. 1, pp. 132 – 148, 2013, special Section: Pervasive Sustainability.

[40] 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*, 2012, pp. 225–238.

[41] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Zomaya, "Network-assisted offloading for mobile cloud applications," in *IEEE International Conference on Communications (ICC)*, June 2015, pp. 5833–5838.

[42] A. Garcia-Saavedra, P. Serrano, A. Banchs, and G. Bianchi, "Energy consumption anatomy of 802.11 devices and its implication on modeling and design," in *8th ACM International Conference on Emerging Networking Experiments and Technologies*, ser. CoNEXT, 2012, pp. 169–180.

[43] G. Foddis, R. Garroppo, S. Giordano, G. Procissi, 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.

[44] B.-J. Kwak, N.-O. Song, and L. E. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Transactions on Networking*, vol. 13, no. 2, pp. 343–355, 2005.

[45] A. Bujari, B. Licar, and C. E. Palazzi, "Movement pattern recognition through smartphone's accelerometer," in *IEEE Consumer communications and networking conference (CCNC)*, 2012, pp. 502–506.

[46] P. Siirtola and J. Röning, "Recognizing human activities user-independently on smartphones based on accelerometer data," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, no. 5, 2012.

[47] 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.

[48] "FXOS8700CQ: Digital Sensor - 3D Accelerometer + 3D Magnetometer," http://cache.nxp.com/files/sensors/doc/data_sheet/FXOS8700CQ.pdf?pspll=1, 2015, Accessed March 20, 2016.

[49] "BMP280, Barometric Pressure Sensors," https://www.bosch-sensortec.com/bst/products/all_products/bmp280, 2015, Accessed March 20, 2016.

[50] 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.

Andrea Capponi (S'16) is a Ph.D. student at the University of Luxembourg. He received the Bachelor Degree in Telecommunication Engineering and the Master Degree in Telecommunication Engineering both from the University of Pisa, Italy. His primary research interests are Mobile Crowdsensing, Internet of Things and Smart Cities.

Claudio Fiandrino (S'14) is a postdoctoral researcher at IMDEA Networks Institute, Madrid, Spain. Claudio obtained his Ph.D. degree at the University of Luxembourg working on the ECO-CLOUD project focusing on energy efficient communications in cloud, mobile cloud and fog computing. He received the Bachelor Degree in Ingegneria Telematica in 2010 and the Master Degree in Computer and Communication Networks Engineering in 2012 both from Politecnico di Torino. Claudio's work on indoor localization over fog computing platforms received the Best Paper Award in IEEE CloudNet 2016. Claudio was a Visiting Ph.D. Student for three months at Clarkson University, NY, USA. He served as Publication and Web Chair at the IEEE International Conference on Cloud Networking (CloudNet 2014) and as TPC member in IEEE CloudNet 2014, IEEE ICC 2016 - SAC Cloud Communications and Networking and IEEE EmergiTech 2016. His primary research interests include mobile cloud/fog computing, mobile crowdsensing and data center communication systems.

Dzmitry Kliazovich (M'03-SM'12) is a Head of Innovation at ExaMotive. He was a Senior Scientist at the Faculty of Science, Technology, and Communication of the University of Luxembourg. Dr. Kliazovich holds an award-winning Ph.D. in Information and Telecommunication Technologies from the University of Trento (Italy). Dr. Kliazovich is a holder of several scientific awards from the IEEE Communications Society and European Research Consortium for Informatics and Mathematics (ERCIM). His works on cloud computing, energy-efficiency, indoor localization, and mobile networks received IEEE/ACM Best Paper Awards. He coordinated organization and chaired a number of highly ranked international conferences and symposia, including the IEEE International Conference on Cloud Networking (CloudNet 2014). Dr. Kliazovich is the author of more than 100 research papers. He is the Associate Editor of the IEEE Communications Surveys and Tutorials and of the IEEE Transactions of Cloud Computing journals. He is a Vice Chair of the IEEE ComSoc Technical Committee on Communications Systems Integration and Modeling. Dr. Kliazovich is a coordinator and principal investigator of the Energy-Efficient Cloud Computing and Communications initiative funded by the National Research Fund of Luxembourg. His main research activities are in the field of intelligent transportation systems, telecommunications, cloud computing, and Internet of Things (IoT).

Pascal Bouvry is a professor in the Computer Science and Communication research unit of the Faculty of Science, Technology and Communication at the University of Luxembourg and a faculty member at the Luxembourg Interdisciplinary Center of Security, Reliability, and Trust. His research interests include cloud & parallel computing, optimization, security and reliability. Prof. Bouvry has a Ph.D. in computer science from the University of Grenoble (INPG), France. He is on the IEEE Cloud Computing and Elsevier Swarm and Evolutionary Computation editorial boards. He is also acting as communication vice-chair of the IEEE STC on Sustainable Computing and co-founder of the IEEE TC on Cybernetics for Cyber-Physical Systems. A full biography is available on page http://pascal.bouvry.org. Contact him at pascal.bouvry@uni.lu.

Stefano Giordano (SM'10) Stefano Giordano received the master's degree in electronics engineering and the Ph.D. degree in information engineering from the Università di Pisa, Pisa, Italy, in 1990 and 1994, respectively. He is currently an Associate Professor with the Department of Information Engineering, Università di Pisa, where he is responsible for the telecommunication networks laboratories. His research interests are telecommunication networks analysis and design, simulation of communication networks, and multimedia communications. He is a member of the Editorial Board of the IEEE COMMUNICATION SURVEYS AND TUTORIALS. He was the Chair of the Communication Systems Integration and Modeling Technical Committee. He is an Associate Editor of the International Journal on Communication Systems and the Journal of Communication Software and Systems technically cosponsored by the IEEE Communication Society. He was one of the referees of the European Union, the National Science Foundation, and the Italian Ministry of Economic Development.

2377-3782 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.