

Sociability-Driven Framework for Data Acquisition in Mobile Crowdsensing over Fog Computing Platforms for Smart Cities

Claudio Fiandrino, *Student Member, IEEE*, Fazel Anjomshoa, *Student Member, IEEE*, Burak Kantarci, *Senior Member, IEEE*, Dzmityr Kliazovich, *Senior Member, IEEE*, Pascal Bouvry, *Member, IEEE*, and Jeanna Matthews, *Member, IEEE*

Abstract—Smart cities exploit the most advanced information technologies to improve and add value to existing public services. Having citizens involved in the process through mobile crowdsensing (MCS) augments the capabilities of the platform without enquiring additional costs. In this paper, we propose a novel framework for data acquisition in MCS deployed over a fog computing platform which facilitates a number of key operations including user recruitment and task completion. Proper data acquisition minimizes the monetary expenditure the platform sustains to recruit and compensate users as well as the energy they spend to sense and deliver data. We propose a new user recruitment policy called DSE (Distance, Sociability, Energy). This policy exploits three criteria: (i) spatial distance between users and tasks, (ii) user sociability, which is an estimate of the willingness of users to contribute to sensing tasks, and (iii) remaining battery charge of the devices. Performance evaluation is conducted in a real urban environment for a large number of participants with new metrics assessing the efficiency of recruitment and the accuracy of task completion. Results reveal that the average number of recruited users improves by nearly 20% if compared to policies using only spatial distance as selection criterion.

Index Terms—Data acquisition, fog computing, internet of things, mobile crowdsensing, smart city sensing, sustainability.

1 INTRODUCTION

WORLD population living in cities has experienced an unprecedented growth over the past century. While only 10% of the population lived in cities during 1900, nowadays this percentage corresponds to 50% and is projected to further increase [1]. Sustainable development plays therefore a crucial role in city development. While only 2% of the world's surface is occupied by urban environments, cities contribute to 80% of global gas emission, 75% of global energy consumption [2] and 60% of residential water use [1]. Street lighting attributes nearly 19% of the worldwide use of electrical energy and entails 6% of global emissions of greenhouse gases [3].

Smart cities rely on Information and Communication Technology (ICT) solutions to improve citizens' quality of life [4], [5]. The application of the Internet of Things (IoT) paradigm to urban scenarios is of special interest to support the smart city vision [5], [6], [7]. IoT is envisioned as the can-

didate building block to develop sustainable ICT platforms. With IoT, everyday life objects are uniquely identifiable and "smart", i.e., they are equipped with computing, storage and sensing capabilities and can communicate one with each other and with the users to enable pervasive and ubiquitous computing [8]. Cisco has recently proposed *fog computing* as an extension of the cloud computing paradigm at the edge of the network [9]. Fog computing is tailored to serve applications that are geodistributed, require low latency and context awareness [10], such as indoor localization [11]. Including citizens in the loop with crowdsensing approaches augments capabilities of existing infrastructures without introducing additional costs and has been proved to be a win-win strategy for smart city applications [12], [13].

Mobile crowdsensing (MCS) has emerged in the recent years, becoming an appealing paradigm for sensing data [14]. In MCS, users contribute data generated from sensors embedded in mobile devices including smartphones, tablets and IoT devices like wearables. The aggregated information is then delivered to a collector. The pervasive diffusion of smartphones and wearables along with the rich set of built-in sensors mobile and IoT devices are equipped with lead to the success of MCS paradigm. Accelerometer, gyroscope, GPS, microphone and camera are a representative set of sensors which are essential to operate a number of applications in health care, environmental and traffic monitoring and management [15]. Google, for example, uses crowd-sourced information about smartphone locations to offer real-time view of congested traffic on roads and has recently released a new application, called Science Journal, which permits visualization of data collected smartphones [16].

- C. Fiandrino is with Imdea Networks Institute, Madrid, Spain. E-mail: {claudio.fiandrino}@imdea.org. Claudio developed the current work as a Ph.D. student at the University of Luxembourg.
- F. Anjomshoa is with the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY, 13699 USA. E-mail: {anjomsm}@clarkson.edu
- B. Kantarci is with the School of Electrical Engineering and Computer Science, University of Ottawa, Ottawa, ON, Canada. E-mail: {burak.kantarci}@uOttawa.ca
- D. Kliazovich is with ExaMotive, Luxembourg. E-mail: kliazovich@ieee.org.
- P. Bouvry are with the University of Luxembourg, Luxembourg. E-mail: pascal.bouvry@uni.lu
- J. Matthews is Department of Computer Science, Clarkson University, Potsdam, NY, 13699 USA. E-mail: jnm@clarkson.edu

In MCS, data acquisition, also known as data collection, can be *participatory* or *opportunistic* [15]. In opportunistic sensing systems, the user involvement is minimal: sensing decisions are application- or device-driven. In participatory sensing systems, users are actively engaged in the sensing process. The users are recruited by a central platform, which dispatches sensing tasks. Users can then decide which request to accept and, after accepting, they have to accomplish specified sensing and data reporting tasks. On one side, opportunistic sensing lowers the burden of user participation as devices or applications are responsible to take sensing decisions. Conversely, participatory sensing systems are tailored to crowdsensing architectures with a “central platform”, which facilitates system control operations like task assignment, user incentives and rewarding to compensate the participants for their contribution.

User recruitment is one of the key challenges in participatory MCS systems. In urban environments, the high number of potential contributors calls for the design of efficient recruitment policies. Implementing proper policies allows selecting users able to fulfill sensing tasks with high accuracy and at minimum costs for the system. From the standpoint of the central platform, which organizes and dispatches tasks, the efficiency of a data acquisition framework is defined in terms of the revenues and the costs sustained. The platform earns revenues through Sensing as a Service (S²aaS) business models [17]. The costs are twofold. The central platform sustains monetary costs to recruit and reward users for their contribution. Users as well sustain costs while contributing data, i.e., the energy spent from the batteries for sensing and reporting data and, eventually, the data subscription plan if the cellular connectivity is utilized for reporting. As in cloud-based S²aaS the mobile devices are the most energy-hungry components in the ecosystem [18], cost-effective solutions in data acquisition not only allow minimizing the energy expenditure, but are also a powerful incentive to stimulate user participation [19].

In this paper, we propose a novel framework for data acquisition in participatory MCS systems. The framework is deployed over a fog computing platform specifically developed for smart cities. The fog facilitates the most important operations in data acquisition, such as user recruitment and task completion. Specifically, we introduce a new user recruitment policy called DSE, which is based on three criteria: (i) the spatial *Distance* between users and tasks, (ii) user *Sociability*, which is an estimation of the willingness of users to participate in and contribute to sensing tasks, and (iii) the *Energy*, computed as the remaining battery charge of user devices. Cloudlets in the fog platform are responsible to estimate sociability, which determines whether users are eligible to contribute data. Highly sociable users share many interests with friends and are more active, i.e., they constantly use devices online, which makes them excellent candidates for data acquisition. Furthermore, we propose novel metrics to assess the efficiency of any recruitment policy and the accuracy of task completion. The User Recruitment Effectiveness (URE) metric evaluates the number of users contacted versus recruited. The Global Task Accuracy (GTA) metric quantifies the accuracy of accomplished task by evaluating the distribution of the data the users contribute along time. Performance evaluation, conducted in a real

urban environment for a large number of participants, reveals the effectiveness of the proposed user recruitment policy as the average number of recruited users improves by nearly 20% if compared to policies using spatial distance as the only selection criterion.

The synopsis of contributions of this work is as follows:

- Propose a novel fog computing platform, which brings computing intelligence close to the end users in a distributed fashion across the city, specifically by using fog computing cloudlets, which are deployed on bus stops throughout the city.
- Develop new data acquisition framework for MCS systems, which exploits the proposed fog platform for user recruitment and task completion.
- Introduce novel metrics to assess efficiency of user recruitment and accuracy of task completion.
- Design of a custom simulator for user recruitment and data acquisition in a large-scale urban scenario.
- Provide performance evaluation of the proposed framework through simulations.

The rest of the paper is organized as follows. Section 2 presents background on MCS data acquisition frameworks and user recruitment policies, motivating the need for social-based and energy-efficient recruitment policies. Section 3 proposes a novel fog computing platform, where the computing capacity is efficiently distributed across the urban environment. Section 4 details the proposed methodology for data acquisition and user recruitment. Section 5 analyzes and evaluates user sociability. Section 6 provides performance evaluation and Section 7 concludes the work outlining future research directions on the topic.

2 BACKGROUND AND MOTIVATION

This section reviews the research in the field of mobile crowdsensing with a focus on participatory data acquisition frameworks and user recruitment policies.

MCS data acquisition platforms are systems in which users contribute information from IoT mobile devices. Such information is then delivered to a collector, typically located in the cloud, to be at disposal of the organizer of the sensing campaign for processing and analysis. Fig. 1 illustrates the main elements of a typical MCS system.

2.1 Data Acquisition in MCS

The main goal of the data acquisition framework is to collect information efficiently. For this, the framework defines a set of steps necessary to produce and deliver information to the collector. Data collection frameworks can be *general-purpose* or *application-specific* [20]. Application-specific frameworks are designed to serve only one type of application at a time. GasMobile [21] and NoiseMap [22] are examples of such data collection frameworks developed to monitor air and noise pollution respectively.

In this paper, we propose a general-purpose framework. Unlike application-specific frameworks, the salient feature of general purpose frameworks is the capability of serving many applications at the same time. BLISS [23] uses an online learning algorithm for general-purpose data collection. The collector optimally assigns tasks to the users having a limited

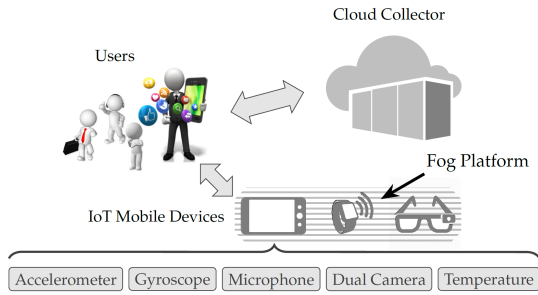


Fig. 1. System scenario

budget available at disposal for rewarding. Wang et al. [24] propose an energy-efficient algorithm for uploading the sensed data. The algorithm groups users into two categories: those who have paid data plan with mobile operators and try to minimize the energy cost during data uploading, and those who want to minimize cost of data uploading benefiting from free-of-charge networks, such as WiFi or Bluetooth. With Piggyback CrowdSensing [19] data uploading can be performed during voice calls. Liu et al. [25] define a new routing mechanism for data collection in MCS systems to cope with user selfishness. Data delivery is always performed through opportunistic communications and is relayed in a delay-tolerant fashion only by non-selfish nodes that are willing to cooperate. To balance the workload among the participants while maximizing the utility of data collection, a Nash bargaining approach is proposed in [26] with two cooperative players, one for load distribution and one for utility maximization.

The closest framework to our work is CARDAP [27]. It extends and augments functionalities of CAROMM [28] by enabling efficient data analytics performed in a distributed fashion in the fog platform. Unlike CARDAP, in this work we exploit the computing capacity provided by the fog platform not for data analytics, but for efficient user recruitment and task completion in participatory MCS systems.

2.2 User Recruitment in MCS

To organize a MCS campaign, the organizer, such as a government agency, an academic institution or a business corporation, sustains costs to recruit and compensate users for their involvement. Therefore, devising efficient recruitment policies is essential. On one hand, it allows organizers to minimize the expenditure. On the other hand, it helps to choose the users that will carry out the campaign successfully. For example, in the context of public safety, it is essential to select users that maximize the trustworthiness of collected data [29], [30], [31], [32], [33].

Several research works investigate task assignment and user recruitment in MCS systems. The majority of the proposed policies aims at minimizing the cost of sensing for organizers while guaranteeing a certain level of system accuracy, such as coverage of the sensing area [34], [35], [36], [37]. Reddy et al. [34] propose a recruitment policy which selects the participants on the basis of their availability for data collection in a given geographical area and at a defined time. In the context of opportunistic sensing systems, Karaliopoulos et al. formulate an optimization problem to minimize system costs and predict user location using deterministic and stochastic mobility models [35]. Hassani

et al. [38] propose Context-Aware Task Allocation (CATA), which is a framework for opportunistic MCS systems allocating tasks to users recruited through a special policy. The policy aims at selecting the most appropriate users for sensing by computing a set of indicators that determine the similarity between the participants and the tasks. Moreover, the policy also considers energy consumption of operations like sensing and data delivery and protects user privacy, e.g., the platform never asks users to reveal sensible information such as user location. Liu et al. [39] propose an energy-efficient participant selection scheme, which relates residual battery charge of user devices to the willingness to contribute. The scheme ensures the quality of the sensed information in terms of the amount of collected data per task. With a comparative analysis of delay-tolerant routing protocols for reporting, Tuncay et al. [40] propose a comprehensive framework that simultaneously addresses both participant recruitment and data acquisition. As users might be disclosing their location during the selection process, in [41] the authors investigate spatial cloaking, which is a promising and effective solution to guarantee privacy.

The closest user recruitment policy to our proposal exploits social relationships to establish a trusted route between service requester and provider parties [42]. More specifically, the service requester is interested in acquiring information on a given phenomenon. If the service provider belongs to the same community of the requester, it receives immediately the sensing task. Otherwise, the task is offered to users belonging to overlapping communities until it is delivered to the service provider. The trust of passages among the communities is guaranteed by social ties between the users within each community.

3 FOG COMPUTING PLATFORM IN SMART CITIES

Fog computing architectures are heterogeneous and include both devices at the edge of the network and traditional cloud data centers. In the front-end, the mobile devices are IoT devices, smartphones and cloudlets with variable amount of computing, storage and networking resources. Cloudlets are typically micro servers or local processing units such as notebook or desktop computers used for temporary storage and processing [43]. In addition to computing resources, they provide data aggregation functions to reduce the amount of information delivered to the cloud. Cloud data centers, located in the back-end of the architecture, provide centralization of functionalities and backup.

To support IoT-based services and applications including MCS in urban environments efficiently, the deployment of fog architectures is needed [44]. Edge devices such as cloudlets, that are responsible for provisioning of location-aware and low latency computing, have to be geographically distributed across the city. Bus stops are a natural choice for installing cloudlets for a number of reasons. First, bus stops are widespread across urban environments. Second, they are already existing and deployed infrastructure. As a consequence, such solution would save capital expenditure costs. For example, bus stops are already connected to the city power grid, therefore there is no need for investments. Third, bus-stops are expected to become *intelligent* in the near future and provide additional community services. For example,

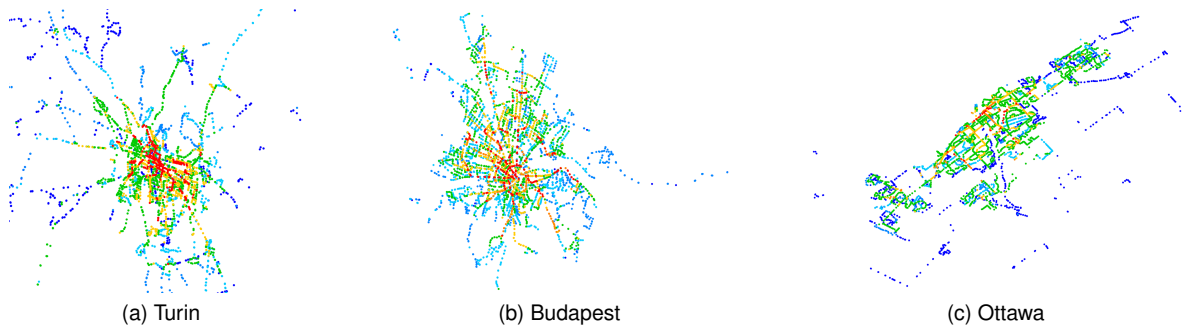


Fig. 2. Analysis of stop relevance distribution in the cities of Turin, Budapest and Ottawa

TABLE 1

Classification of stop relevance. Each interval corresponds to the number of trips passing by a stop from 6:00 AM to 10:00 PM.

COLOR	RELEVANCE INTERVAL		
	TORINO	BUDAPEST	OTTAWA
●	0 - 5	0 - 25	0 - 25
●	5 - 10	25 - 100	25 - 50
●	10 - 50	100 - 250	50 - 100
●	50 - 100	250 - 500	100 - 250
●	100 - 250	500 - 750	250 - 500
●	250 - 500	750 - 1000	500 - 1000
●	>500	>1000	>1000

but stops will host femto cells to increase cellular capacity and connectivity [45]. In cities, bus stops are not all identical. Because of the location, some of them are hubs, i.e., they are stops used by a huge number of passengers every day. Therefore, such stops are more *relevant* to the system than the ones used by few passengers. Obviously, stops need to be equipped with computing capacity proportionally to their relevance. To study the relevance property of bus stops, we analyze data from Google Transit Feed. Google Transit [46] is a tool that integrates information on public transportation system like stop location, routes, bus schedule and fare on Google Maps to let trip planning easier and accessible for everyone.

Similarly to [47], the *relevance* of stop i is defined in terms of the number of trips crossing i during a given time period. For the evaluation, the time period is set to 6:00 AM - 10:00 PM of a weekday, namely 2016-06-17. Fig. 2 shows relevance stop distribution in different cities, namely Torino, Budapest and Ottawa according to the classification proposed in Table 1. In all the cities, the number of stops with high relevance is low. Relevant stops are typically concentrated in the city center or in proximity of popular public parking facilities or train stations, which are expected to serve a high number of users along the day.

4 DATA ACQUISITION FRAMEWORK

A data acquisition framework defines the set of steps necessary to produce and deliver the information from the participants to the organizer/collector. Fig. 3 shows the architecture considered and illustrates the functions each actor carries out.

The *organizer* of the crowdsensing campaign \mathcal{C} is interested in acquiring data from a set of *points of interest* in the city, also called the *sensing terrain*. The organizer, located in the cloud, is in charge of analyzing data after it has been collected

and make it available to S^2aaS applications. The organizer defines the set of sensing tasks $\mathcal{W} = \{w_1, w_2, \dots, w_W\}$ of \mathcal{C} . Each task w_i is described in terms of its location L_i and time duration T_i , i.e., $w_i(L_i, T_i)$. The location L consists of latitude and longitude parameters, defining the center of the area of interest. The time duration T is given in timeslots. As a result, the duration \mathcal{T} of the campaign \mathcal{C} is as follows:

$$\mathcal{T} = \sum_{i \in \mathcal{W}} T_i. \quad (1)$$

In this work, the user recruitment policy exploits the computing capacity of the fog platform to determine user eligibility. It is worth noting that user recruitment is not the sole application that the fog platform deployed as in Section 3 can support. Local analytics [27], [48], tag affinity [49], privacy preservation and evaluation of trust of contributed data [50], [51], [52] are a few examples of functionalities that the proposed fog platform can support. The cloudlets receive the tasks to dispatch from the organizer in the cloud. The cloudlets in proximity of the location of the sensing tasks are in charge of recruiting the users with the policy proposed in Section 4.1. The participants communicate periodically to the cloudlets information about their sociability and remaining battery charge. Consequently, the cloudlets can determine which users are eligible to become contributors and contact them for the assignment.

After being contacted, the users can decide whether to accept the task. In positive case, users acquire the status of *recruited*, they are *assigned* to the task and can contribute data. Acceptance depends on user sociability and remaining battery of charge of user devices. Users with high values of sociability factor use social media more often and are likely to accept the task [53].

4.1 User Recruitment

User recruitment is a fundamental step in participatory data acquisition frameworks. Recruitment policies delineate the set of criteria for user eligibility in contributing to crowdsensing campaign. Contrary to traditional recruitment solutions, in this paper we define a policy able to select participants on the basis of three parameters: (i) the distance between users and sensing task location (D), (ii) user sociability (S), and (iii) remaining battery of charge of users devices (E). The policy is named DSE (Distance, Sociability, Energy). The parameters D , S and E_i are unit-less and can assume real values in $[0, 1]$. Table 2 lists description of symbols used in the model.

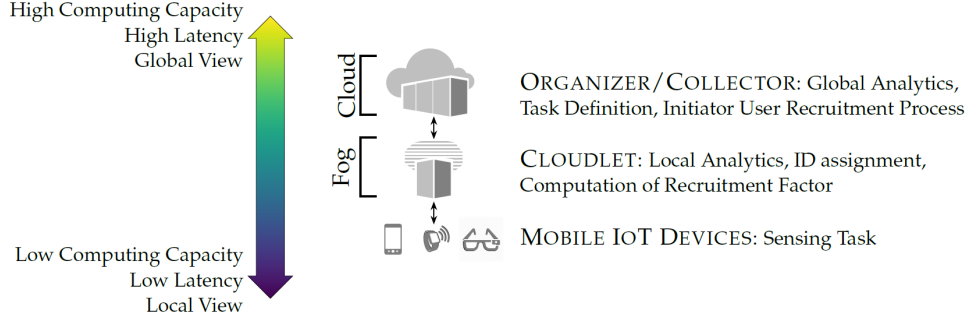


Fig. 3. Architecture of the proposed fog computing-based data acquisition framework

TABLE 2
Description of symbols

SYMBOL	DESCRIPTION
\mathcal{C}	Crowdsensing campaign
w	Sensing task w
\mathcal{W}	Set of tasks $ w \in \mathcal{W}$
u	User u
\mathcal{U}	Set of users $ u \in \mathcal{U}$
t	Timeslot t
T_i	Duration of task i
\mathcal{T}	Duration of the sensing campaign
L	Location of users and tasks
R_i	Recruitment factor of user i
D_i	Distance factor of user i
S_i	Sociability factor of user i
E_i	Energy factor of user i
R_{\min}	Minimum recruitment factor for eligibility
$D_{u,w}$	Distance (m) between user u and task w
D_{\max}	Maximum task coverage radius
E_s	Energy consumed for sensing
E_r	Energy consumed for data delivery (reporting)
P_{tx}	Power consumed during data delivery
A_i	Task acceptance factor of user i
N	Minimum number of users to mark a task as <i>accomplished</i>
P_i	Popularity factor of location i
$\mathbb{E}[N^c]$	Average number of contacted users
$\mathbb{E}[N^r]$	Average number of recruited users

Let $\mathcal{U} = \{u_1, u_2, \dots, u_U\}$ be the set of users potentially available to perform sensing. Each user u_i is described in terms of his/her current location, sociability factor and energy, i.e. $u_i(L_i, S_i, E_i)$. Both user location and sociability factor are time dependent.

During each timeslot, the recruitment policy selects users with highest *recruitment factor* R from the set \mathcal{U} . Only users with values of $R > R_{\min}$ are taken into consideration and *contacted*. R_{\min} defines the minimum recruitment factor and is set by the organizer to be identical for all the tasks in the campaign \mathcal{C} . For each user i , the *recruitment factor* is defined as follows:

$$R_i = \alpha \cdot D_i + \beta \cdot S_i + \gamma \cdot E_i, \quad (2)$$

where the parameters α, β, γ are weighted coefficients defining the impact of the corresponding component, distance, sociability and energy on R . Taking into account that $\alpha + \beta + \gamma$ must equal unity, high values of α will prioritize selection of users close to the location of sensing tasks. High values of β will favor selection of highly sociable users while high values of γ will make the remaining battery charge of devices the most important component for recruitment. Section 6.3.3 details how to properly set the parameters to maximize the number of the successfully accomplished tasks.

The component D_i is the distance factor, which measures the distance of user u_i from the location of sensing task w_j with respect to a maximum coverage radius for the task D_{\max} .

$$D_i = 1 - (D_{u_i, w_j} / D_{\max}). \quad (3)$$

Users located farther than D_{\max} from the location of a sensing task are not considered eligible to contribute data for that task. Indeed, the closer the users are to the sensing task location, the higher the accuracy in capturing the phenomenon is. The Haversine formula can be employed to compute D_{u_i, w_j} [54].

User sociability S can be defined in terms of the amount of data users consume or the time they spend using mobile social network applications, or their combination [55]. Sociability is an essential parameter to consider for user recruitment. Users with high sociability are more active and use their devices online intensively, which makes them excellent candidates during the selection process. Moreover, they tend to visit more places and get connected to more users, which further increases their mobile social activity [56]. Section 5 provides more information and deep analysis on how to compute user sociability.

The parameter E indicates the energy, i.e., the remaining battery charge of user devices. E is directly measured from the mobile or IoT device operating system and normalized to be in the range $[0, 1]$. Indeed, the most widely adopted mobile OS like iOS and Android provide APIs to obtain information on current level of battery charge expressed in percentage. For crowdsensing operation, the devices consume energy to perform sensing (E_s) and reporting (E_r) operations:

$$E = E_s + E_r. \quad (4)$$

The energy E_s drain due to sensing is the sum over all sensors \mathcal{K} involved to fulfill a task during T :

$$\sum_{k \in \mathcal{K}} \sum_{t=1}^T f_k \cdot \rho_k, \quad (5)$$

where f_k is the sampling frequency of sensor k and ρ_k a constant, different per sensor, which describes the energy cost per sample [27]. Typically, the parameter ρ can be obtained from the data sheet of the sensor.

The users exploit WiFi connectivity for data reporting and communication with the cloudlets. Most of the mobile operating systems, including Android and iOS, tend to prefer WiFi over cellular connectivity for data transmission, as it is more energy efficient [57] and users do not consume the data plan they pay to the cellular operators [58]. As a result, when both WiFi and LTE interfaces are active, transmissions take

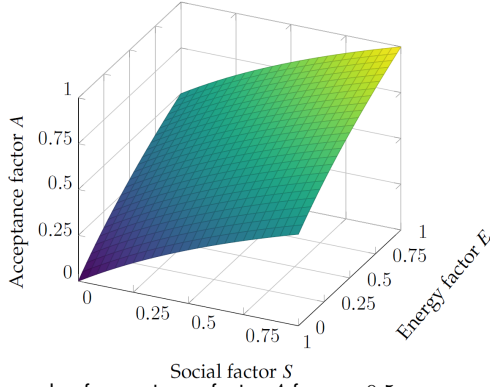


Fig. 4. Example of acceptance factor A for $\sigma = 0.5$

place via WiFi. The energy E_r spent during the transmission time τ_{tx} is defined as:

$$E_r = \int_0^{\tau_{tx}} P_{tx} dt, \quad (6)$$

where P_{tx} is the power consumed for transmissions of WiFi packets generated at rate λ_g [59]:

$$P_{tx} = \rho_{id} + \rho_{tx} \cdot \tau_{tx} + \gamma_{xg} \cdot \lambda_g. \quad (7)$$

4.2 Task Completion

To recruit users, the campaign organizer sustains a cost. For each request sent to the users, the cost c associated to the task w is equal to 1 unit of cost. The costs have a different nature. For example, costs could be financial or expressed in terms of the bandwidth used to broadcast recruitment messages. Costs can also be social: contacting persistently a users who has refused to accomplish to a task in previous timeslots, it is likely to diminish the chances that s/he will accept the task. The objective of the organizer is to minimize the total cost sustained while maximizing the number of accomplished tasks. The tradeoff between the recruitment cost and the number of accomplished tasks defines the efficiency of the recruitment policy.

Users with high recruitment factor are contacted and can decide whether to accept or refuse the task. Upon acceptance, the user acquires the *recruited status*. Acceptance is based on user sociability and remaining battery charge. Users with high values of sociability and energy factors, S and E respectively, are more likely to accept the task. The acceptance factor A is computed by the user devices and it is modelled as a logarithmically increasing function:

$$A(S, E) = \sigma \cdot \log(1 + S) + (1 - \sigma) \cdot \log(1 + E), \quad (8)$$

where σ is a balancing coefficient that shows a relative importance between the sociability and energy factors. Fig. 4 shows the relation between A , S and E , which allows to perform a fine-grain comparison of the task acceptance probability of users with low versus high sociability and energy ratings. For users with high values of sociability and remaining battery charge, the acceptance factor A assumes values close to 1. Vice versa, for users with low values of sociability and remaining battery charge, a small difference between two factors S_1 and S_2 , E_1 and E_2 corresponds to a considerable difference in the respective acceptance factors A_1 and A_2 .

Upon acceptance, the user acquires the *recruited status* and contributes as long as s/he remains within a distance closer than D_{max} . In such a case, s/he is not contacted to contribute

to the same task any longer. Viceversa, users refusing a task can be contacted again if the eligibility criteria are still met. After rejection during timeslot t , a user is contacted again at timeslot t_{next} , which is defined as follows:

$$t_{next} = t + j \cdot \tau. \quad (9)$$

The parameter τ is a fixed number of timeslots the systems backs off and j is the number of times the user has previously refused the same task. Consequently, the higher the number of rejection, the longer the system will wait before contacting again the user for the same task.

System-level accuracy increases if the organizer does not recruit persistently the same group of users to accomplish a task [60]. For this reason, each task w acquires the status *accomplished* if, during t , a given number N of individual users are involved and contribute by reporting data. During t_i , whenever it is not possible to recruit a sufficient number of users, the task i is marked as *failed*.

Like in social networks, some locations in cities are *hubs*, i.e., they attract a large number of individuals, whereas others do not [61]. To capture this phenomenon, each location l is assigned a popularity factor P , and P can take real values in the range $[0, 1]$. Practically, tasks associated to locations with high popularity factor should require a high number of users to successfully complete the task. In addition to the location popularity, also the time dimension plays a crucial role in defining N . Longer tasks require a higher number of users than short ones to guarantee good levels of accuracy. As a result, the number of users N_i necessary to accomplish the task i out of \mathcal{U} is calculated as follows:

$$N_i = P_i \cdot (t_i/T) \cdot \mathcal{U}. \quad (10)$$

5 ANALYSIS OF USER SOCIABILITY

User sociability is one of the key parameters defining the recruitment factor. Assessment of user sociability concept through smart mobile devices was initially introduced with the TrackMaison (Track My Activity in Social Networks) framework in [55]. This study has been followed by [62], where a framework was developed to continuously track users interactions through popular social network services. Each interaction through the device is translated into a session that keeps track of the duration of the interaction, data usage during the session, and location information to detect possible anomalies. A normalized running average value of the social network data usage of a user indicates their sociability in a pool of mobile users. In the context of crowdsensing, these users form the pool of participants.

TrackMaison analyzes data gathered from smartphone sensors and users' social network interactions for continuous user identification/authentication with online behavioral biometrics, also called *behaviormetric identification/authentication*. The social networks considered are Facebook, LinkedIn, Whatsapp, Skype and Twitter. More in detail, the framework collects information on the location of users, their data usage, the number of sessions and the session duration. TrackMaison monitors the user interaction with social medias, trains the classifier and then feeds the identification process.

Two metrics are defined for identification/authentication: the *sociability activity rate* and the *sociability factor*. The first is based on the amount of data that a user produces when using

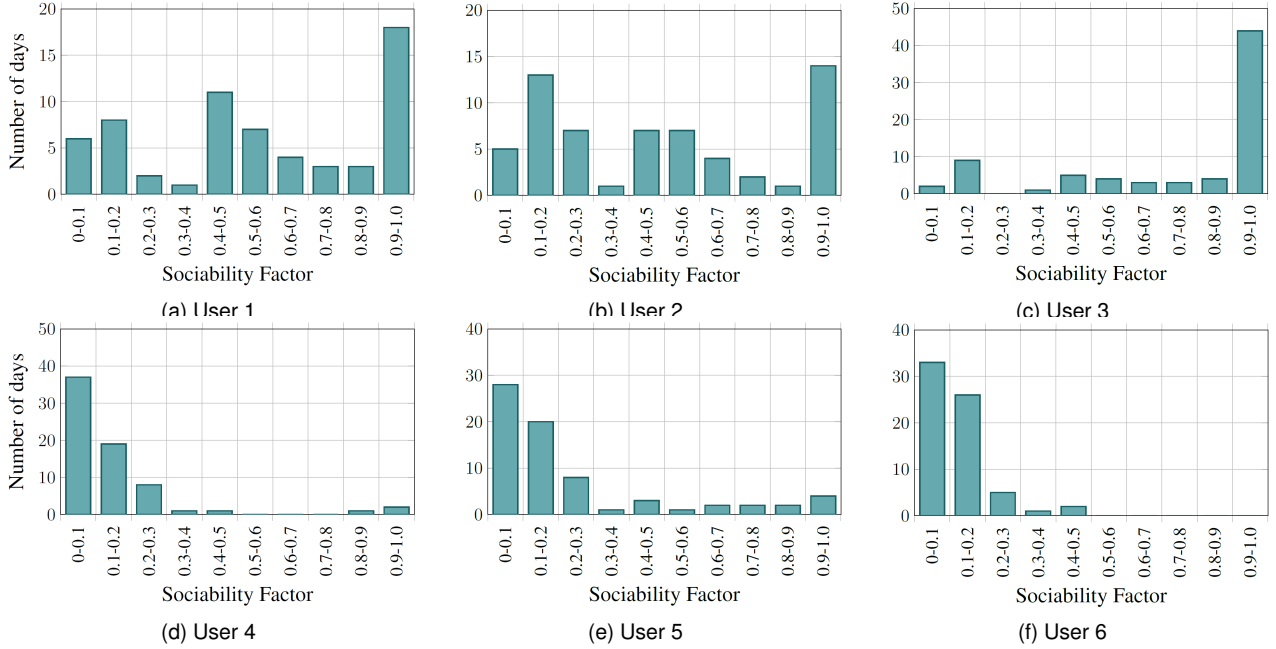


Fig. 5. User profiles for sociability

social networking applications. The amount of time a user spends on social networks defines the second metric, which is used in the paper for user recruitment. The sociability factor can be determined as the instantaneous, the short term or the global component, or overall values. The instantaneous sociability factor \mathcal{SF}_{inssi}^u is calculated as the total time a user spends on a social networking application in a single session.

$$\mathcal{SF}_{inssi}^u = t_i^u. \quad (11)$$

To compute the short term sociability factor \mathcal{SF}_{sh}^u , the time spent on a social network application is averaged over over a short time window τ , e.g., a day.

$$\mathcal{SF}_{sh}^u = \left(\sum t_i^u \right) / \tau. \quad (12)$$

The global component $\mathcal{SF}_{overall}^u(T_k)$ is defined as the weighted sum of short term sociability factors:

$$\mathcal{SF}_{overall}^u(T_k) = \rho \cdot \mathcal{SF}_{sh}^u(T_k) + (1 - \rho) \cdot \mathcal{SF}_{overall}^u(T_{k-1}), \quad (13)$$

where T_k denotes the k -th short term sociability factor used in the calculation, and ρ is a weight factor defining the importance of short term social factor value and previous overall social factor value. From the $\mathcal{SF}_{overall}^u(T_k)$ values, the cloudlets determine S among all the users contributing data to the campaign. For each user, the cloudlets compute S as the aggregated overall sociability factors scaled by the maximum aggregated sociability factors in the active users pool.

$$S = \sum_{x \in \mathcal{X}} \omega_x \mathcal{SF}_{overall}^u(T_k) / \arg \max_{u \in \mathcal{U}} \sum_{x \in \mathcal{X}} \omega_x \mathcal{SF}_{overall}^u. \quad (14)$$

The parameter ω_x is a weight factor unique for each mobile social network application.

Fig. 5 and Table 3 show the analysis of sociability factor performed over a real data set, with information collected

 TABLE 3
 User Sociability Factors

PROFILE	SOCIAL FACTOR S	
	MEAN	STANDARD DEVIATION
User 1	0.472	0.264
User 2	0.408	0.270
User 3	0.754	0.332
User 4	0.133	0.359
User 5	0.234	0.314
User 6	0.100	0.374

from 6 users for a period of 79 days. Each plot represents the number of days each user has achieved a overall social factor. Users are divided into two groups. In the first group, Users 1, 2 and 3, show highly sociable profiles as for the majority of the days they spend lot of time on social networks. The second group of users, Users 4, 5 and 6, show low sociable profiles as they achieve a low sociability factor for the majority of the evaluation period.

6 PERFORMANCE EVALUATION

This section illustrates performance evaluation of the proposed user recruitment policy for data acquisition in mobile crowdsensing systems. We first propose novel performance metrics and illustrate others Key Performance Indicators (KPI) used for evaluation. Second, we outline the research questions that the simulation results answer and, finally, we analyze the simulation results.

6.1 Performance Metrics

The effectiveness of any recruitment policy can be defined in terms of the number of contacted users that are actually recruited. To quantify such effectiveness, we propose a novel metric called User Recruitment Effectiveness (URE):

$$\text{URE} = \frac{\mathbb{E}[N^r]}{\mathbb{E}[N^c]}, \quad (15)$$

TABLE 4
Sensor and communication equipment parameters used for performance evaluation

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

(a) Sensor

SYMBOL	VALUE	UNIT	DESCRIPTION
ρ_{id}	3.68	W	Power 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

(b) Communication

where $\mathbb{E}[N^c]$ and $\mathbb{E}[N^r]$ correspond to the average number of contacted and recruited users respectively. The URE metric can assume real values in the range $[0, 1]$. Values of URE close to 1 indicate efficient policies. Specifically, $\text{URE} = 1$ indicates that all the contacted users were actually recruited.

In addition to measuring effectiveness of user recruitment, it is also important to assess the number of the assigned users per task and the number of accomplished tasks. Having the capability to measure the former KPI allows the organizer to properly set the recruitment policy. For example, relaxing eligibility constraints for tasks with a low number of users assigned would increase participation. Measuring the number of accomplished tasks defines the overall system effectiveness.

Having defined *accomplished* tasks, it becomes necessary to measure their accuracy. Obviously, task accuracy depends on time distribution of the contribution provided by the N users. Let us consider the case for all the N users deliver sensed data during the first timeslot $t = 1$ and remain idle for the remaining $T - 1$ timeslots. Consequently, the task is accomplished with poor accuracy. Conversely, if N users contribute data *uniformly* along the entire period T , the task is accomplished with high accuracy. The Global Task Accuracy (GTA) metric quantifies accuracy of accomplished task as follows:

$$\text{GTA} = \frac{1}{T} \cdot \left[\sum_{t=1}^T x_t \cdot \frac{n_t}{N} - \sum_{t=1}^T q \cdot \left(\sum_{j=1}^t y_t \right) \right], \quad (16)$$

where n_t is the number of users in timeslot t contributing data and q is a penalization term, which reduces task accuracy when in a given timeslot the contribution is null. The term q is set to be inverse proportional to the number of timeslots T :

$$q = 1/T. \quad (17)$$

The rationale behind this choice is that having no contribution in one timeslot affects more severely short tasks than longer ones. The penalization should be more severe if during consecutive timeslots none of the users contributed data. For this reason, in (16), q linearly increases with the number of timeslots with no contribution. The terms x_t and y_t are boolean variables:

$$x_t = \begin{cases} 1 & \text{if } n_t > 0; \\ 0 & \text{otherwise.} \end{cases}, \quad y_t = \begin{cases} 1 & \text{if } n_t = 0; \\ 0 & \text{otherwise.} \end{cases} \quad (18)$$

It is worth mentioning that both URE and GTA metrics should not be computed run time, but after task completion, i.e., when the duration T_i of the task i has expired.

To properly define the accuracy of data collection, it is necessary to measure the accuracy of sensing reading in addition to the accuracy of task completion. The former is application dependent and, for applications aiming at monitoring phenomena like air or noise pollution, recruitment policies play an essential role. Indeed, accuracy in monitoring the phenomena increase if the same users are not persistently recruited to accomplish a task [60].

6.2 Objectives of the Simulations Experiments

Simulation experiments are the candidate tool to assess performance of MCS systems. The high number of participants makes difficult to perform realistic experiments with testbeds. Objective of the simulation experiments is to validate the effectiveness of the sociability-driven recruitment policy. Analysis of the computing, memory and storage requirements of the fog platform is left to future works.

Simulation results seek answer to the following hypothesis. As mobile applications are mostly used for interaction with cloud services, if users can be identified based on social contexts, those with higher social activity can be recruited frequently, and the others can save energy. An interesting trade-off could occur between the URE and energy/battery drain of the users that are identified as highly sociable. The objective of this paper is evaluating the energy consumption and sensing costs of the overall crowd. Thus, while highly sociable users can be preferred for the sake of accuracy and recruitment effectiveness, they should also be compensated effectively due to high communication and sensing costs. However, addressing proper rewarding based on communication and sensing costs is left to future work as this problem falls into the scope of user incentives in crowdsensing. By evaluating the performance of the proposed scheme, answers to the following research questions are sought:

- What is the impact of sociability-driven recruitment policy on the effectiveness of recruiting the contacted users and on the accuracy of the accomplished tasks?
- How are the sensing and communication costs impacted when sociability is a key recruitment criterion?
- How the parameters of the framework impact the main objective, the number of accomplished tasks?

6.3 Simulation Results

To evaluate and assess efficiency of the data acquisition framework, we exploit CrowdSenSim [63]. CrowdSenSim allows the researchers to perform analysis of crowdsensing

TABLE 5
Simulation settings

PARAMETER	VALUE
Number of users	[1 000 - 10 000]
Overall evaluation period	8:00 AM - 2:00 PM
Time of travel per user	Uniformly distributed in [10, 30] min
Average user velocity	Uniformly distributed in [1, 1.5] m/s
Initial remaining charge of the battery	Uniformly distributed in [0.5, 0.9]
Timeslot duration	1 minute
Task duration	40 timeslots
Number of tasks	25
Number of cloudlets	6
Popularity factor P	0.2

activities in realistic urban scenarios. In this work, the experiments take place in the city center of Luxembourg, see Fig. 6(a).

The number of participants ranges from 2 000 to 10 000, which corresponds to nearly one tenth of the population of Luxembourg (107 340 inhabitants as of late 2014). For simplicity, the start time of the walk is uniformly distributed between 8:00 AM and 1:30 PM. Users walk for a period of time that is uniformly distributed in [10, 30] minutes with an average speed uniformly distributed in [1, 1.5] m/s. Each participant has only one mobile device, whose initial remaining charge of the battery is uniformly distributed between [0.5, 0.9]. The devices are equipped with accelerometer, temperature and pressure sensors, and transmit information using WiFi. As sensing equipment, the devices exploit real sensors implemented in current smartphones and tablets. Specifically, we select the FXOS8700CQ 3axis linear accelerometer from Freescale Semiconductor [64] and the BMP280 from Bosch [65], which is a digital pressure and temperature sensor. Equation (7) describes WiFi power consumption the devices spend for communication. Table 4 presents the detailed information on communication and the parameters. The participants push data to the collector while walking. Once the period of walking ends, they stop moving and contributing. As a consequence, users can contribute for only a small portion of the day, which allows us to study the system performance under a relatively worst case scenario. Each user has an associated sociability profile according to Table 3.

A set of 6 cloudlets dispatching 25 tasks is deployed in different locations of the city, see Fig. 6(b). The starting time of each task is distributed uniformly in the time period 8:00 AM - 2:00 PM. Table 5 lists the details on the simulation settings. For simplicity, each task lasts 40 timeslots and each timeslot corresponds to 1 minute. In the first set of experiments, the popularity factor P of each location, the minimum recruitment factor R_{\min} and the maximum task coverage radius D_{\max} are fixed and set equal to 0.2, 0.55 and 55 m respectively.

6.3.1 Performance of the DSE Policy

Having fixed the parameters of the recruitment policy $\alpha = \beta = \gamma = 0.33$ and $\sigma = 0.5$, Fig. 7 shows the number of contacted and recruited or assigned users per task. Tasks are grouped according to the initial time of deployment. The number of contacted users corresponds to the cost the system sustains for recruitment. In this experiment, we compare the performance of the DSE policy with a

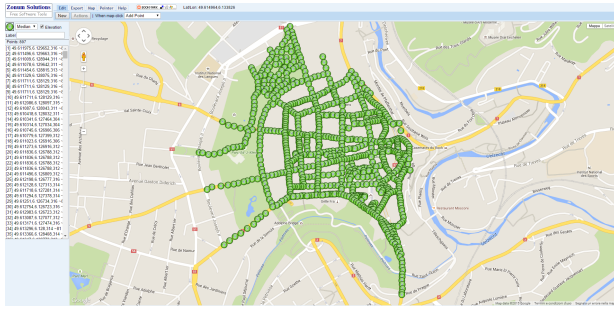
recruitment policy where the distance is the only criterion defining user eligibility. We denote this policy as Distance-Based policy (DB) and requires to set $\alpha = 1, \beta = \gamma = 0$. The DSE policy outperforms the DB policy in terms of the number of users recruited. The average number of recruited users per task is 21.23 and 17.08 for DSE and DB respectively, which corresponds to an increase of 19.55 %. Moreover, when the DB policy fails to contact users like in task # 3 or contacts very few users like in task # 1, the DSE policy makes a significant difference. Consequently, considering user sociability and remaining battery of charge of the devices in addition to task spatial coverage is an effective solution for recruitment. Although being more efficient in recruiting users, the DSE policy is more costly than the DB policy. The average number of contacted users is 45.4, while for the DB policy is 32.92, which corresponds to an increase of costs of around 27 %.

Fig. 8 details the number of unique users assigned to each task. The gray line plots N , the minimum number of users necessary to denote a task as accomplished. N is computed by (10) and is equal for all the tasks as the location popularity and the task duration have been fixed. Consequently, partial relaxation of any of the constraints on task completion would increase the number of accomplished tasks. As it is possible to see, DSE accomplishes 10 tasks out of 25 and one more is close to completion. On the other hand, DB accomplishes 7 tasks and two are close to completion. Only the campaign organizer can compare the tradeoff between cost increase and return, and pursue proper measures, e.g., to reduce the cost of user recruitment.

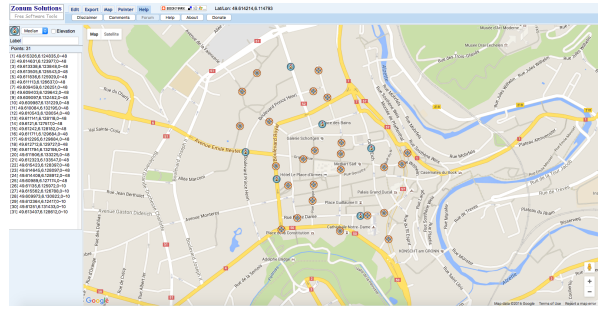
Fig. 9 evaluates the efficiency of recruitment and accuracy of task completion with URE and GTA metrics. The recruitment efficiency of the campaign, computed as an average of the URE values of each task, is 0.46. As a consequence, under the current settings, the systems contacts many users that refuse to contribute to the campaign. It is worth mentioning that for accomplished tasks such as task # 2 (see Fig. 8), the values of URE metric are always equal to or higher than 0.5. Therefore, for accomplished tasks, the recruitment process is more efficient as at least half of the contacted users are actually recruited. The system achieves the highest efficiency in recruitment for task # 3 although the number of recruited users is only 4 and the task is not successfully accomplished. Fig. 9(b) plots values of GTA metric per task. In general, the tasks are not carried out with very high accuracy. The main reason is that only few users out of N contribute to the task. The second most important reason is that not all the N users remain for the entire duration of the task under the maximum task coverage radius D_{\max} . For example, task # 3 is carried out with very low accuracy because few users contribute for a very short duration. As a result the penalization component in (16) diminishes considerably the achieved accuracy.

6.3.2 Analysis of User Energy Consumption

This section evaluates the battery drain of user devices. Fig. 10 shows the distribution of user energy consumption for sensing and data reporting. As expected, both distributions follow the same profile because data after being collected from the sensor is immediately delivered. Fig. 10(a) shows the distribution of users battery consumption due to sensing operations. The results are measured in form of current drain. The vast majority of the users spends little amount of energy

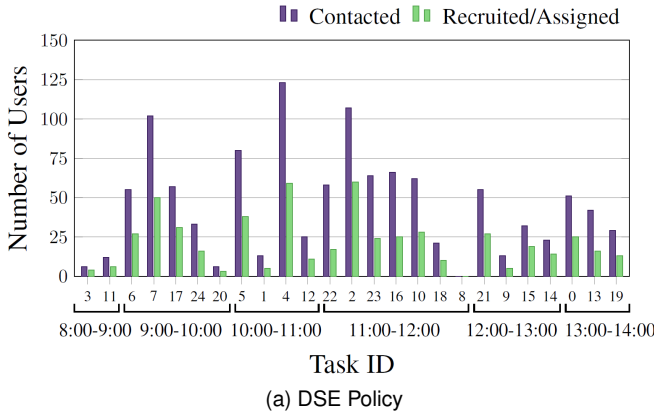


(a) Street-level information

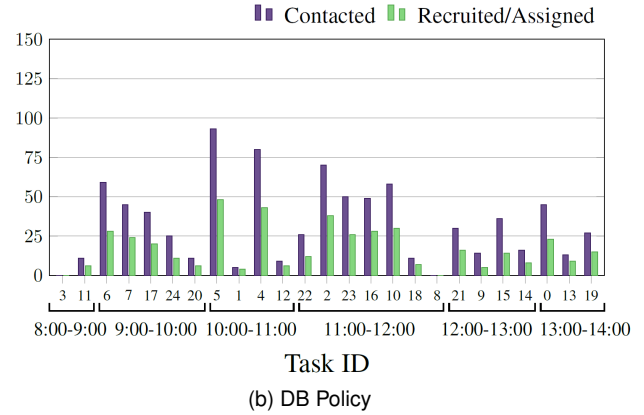


(b) Location of sensing tasks and cloudlets

Fig. 6. Map of Luxembourg



(a) DSE Policy



(b) DB Policy

Fig. 7. Number of contacted versus recruited users. Distance-based (DB) policy is used as baseline for comparison.

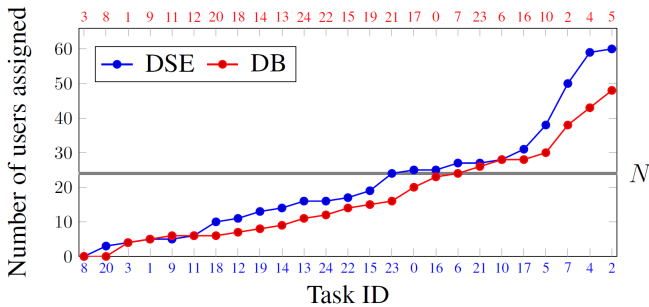


Fig. 8. Number of assigned users per task

for sensing. The motivation is twofold. First, many users contribute to only one task and because of the mobility, they contribute for few timeslots. Second, modern sensors are designed to be energy efficient. When compared with the battery capacity available in today smartphones, which is in the order of 2500 mAh, it is clear that the energy consumed for sensing is negligible with respect to the energy spent for communications (see Fig. 10(b)).

6.3.3 Analysis of Model Control Parameters

Having evaluated the performance of the DSE policy, in this section we study the impact of all main parameters such as the minimum recruitment factor R_{min} and the maximum task coverage radius D_{max} . We also evaluate the influence of the number of users presents in the system as well as the control parameters α , β , γ and σ . For the evaluation, we exploit as common performance metric the average, over 100 runs, of the number of successfully accomplished tasks. The results are expressed in percentage and the bars indicate the 95% confidence interval.

Fig. 11 evaluates the impact of the minimum recruitment factor R_{min} . Proper tuning of this parameter is important: high values of R_{min} make the user selection strict and only few users will be eligible for recruitment. On the other hand, low values of R_{min} relax the conditions for eligibility allowing the system to contact more users. The plot confirms the model: for values of $R_{min} < 0.5$, the percentage of accomplished tasks is higher than 10%. Indeed, contacting more users it increases the chances of having exactly N users assigned to a task, which is the minimum number of individual users necessary to denote the task as successfully accomplished.

Having fixed $R_{min} = 0.3$ from this point on, Fig. 12 evaluates the impact of the maximum task coverage radius D_{max} . The higher the values D_{max} assume, the larger is the area users can be contacted. The plot highlights this property and it is interesting to note a linear increase of the percentage of accomplished tasks for values of D_{max} in the range [30 – 70] m, while for $D_{max} > 70$ the increase becomes smoother. The behavior suggests that increasing the maximum task coverage radius significantly helps to contact higher number of users and to cope with user movement. However, recruiting users far from the location of the sensing task may result in poor accuracy for particular applications. For example, if the sensing task requires users to take a picture, being closer to the location of the task is crucial. On the other hand, for the vast majority of S^2aaS applications requiring monitoring of phenomena such as noise or air pollution, temperature and pressure, having users far for the location of the sensing task can be acceptable.

The previous experiments were conducted having fixed

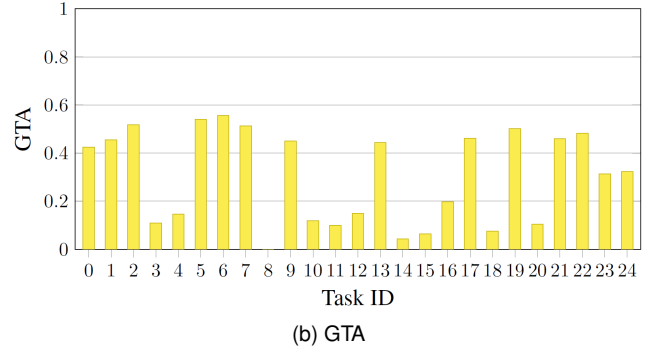
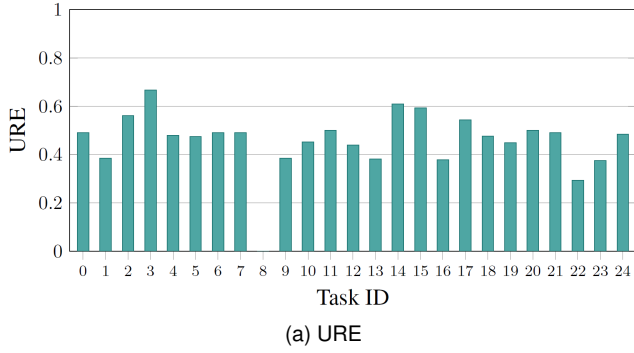


Fig. 9. Evaluation of recruitment efficiency and task accuracy with the new performance metrics

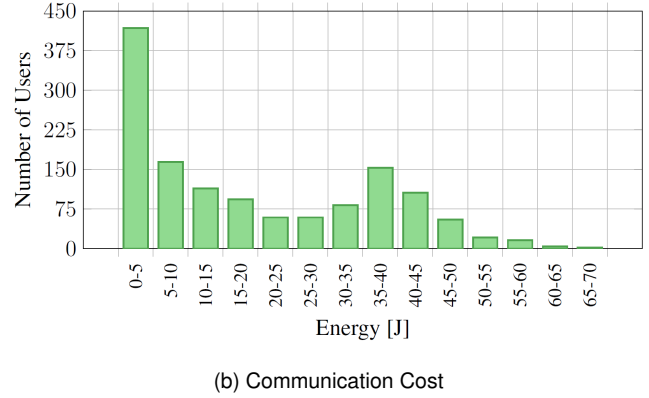
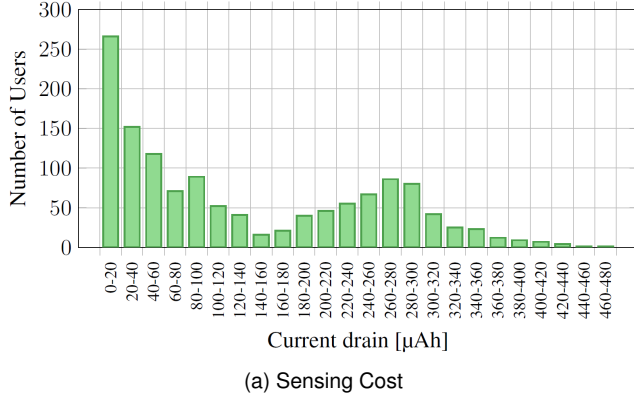


Fig. 10. User distribution for energy

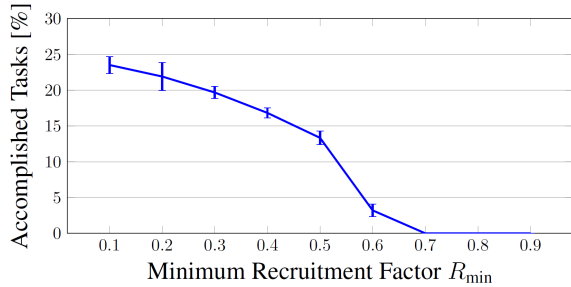


Fig. 11. Average number of successfully accomplished tasks with increasing values of minimum recruitment factor R_{min}

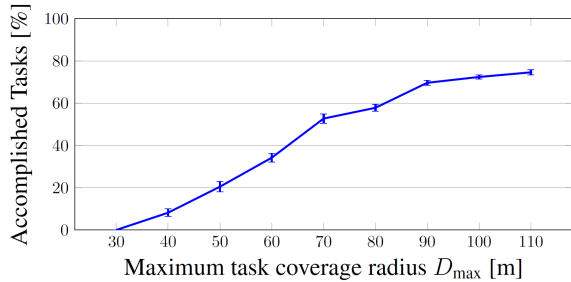


Fig. 12. Average number of successfully accomplished tasks with increasing values of radius D_{max}

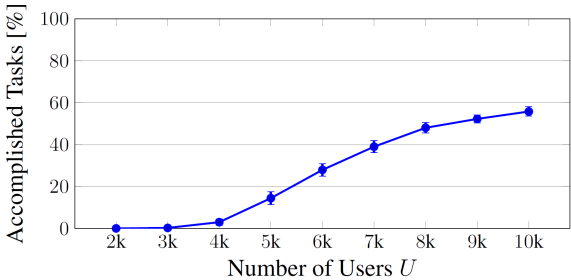


Fig. 13. Average number of successfully accomplished tasks with increasing number of users in the system

the population. The following analysis aims to assess the impact of the total number of users in the system U . A higher number of users in the system makes larger the selection pool and, intuitively, should favor task accomplishment. However, the hypothesis has to be verified for two reasons: (i) the proportionality between N and U , see (10), and (ii) the limited period the users move with respect to the total evaluation period (see Table 5) may lead users be active when the task already expired or did not started yet. Having set $D_{max} = 70$ m, Fig. 13 shows that the number of tasks accomplished increases with the number of users U . The number of accomplished tasks increases substantially (around 40%), when the population in the system changes from 3k to 7k. Because of the aforementioned reasons, the increase becomes more gentle when the number of users in the systems is greater than 7k.

Having analyzed the impact of the population on the system, in the next experiments we verify the sensitivity of DSE to the control parameters α , β and γ . These parameters define the importance of distance, sociability and energy to compute R , see (2). In each experiment, we vary the value of one parameter, while the other two are set identically so that the sum equals one. For the analysis, the number of users in the system is set to 10 000. Fig. 14(a) shows the impact of α . High values of α make the distance the most important factor for recruitment and the plot shows that for $\alpha > 0.4$ the percentage of successfully accomplished tasks remains almost constant. As a result, the selection of any values of $\alpha > 0.4$ are a good choice to maximize the number of accomplished tasks. Fig. 14(b) shows the impact of β , which defines the importance of sociability for R . Unlike the previous case, setting the parameter $\beta = 0.3$ leads to

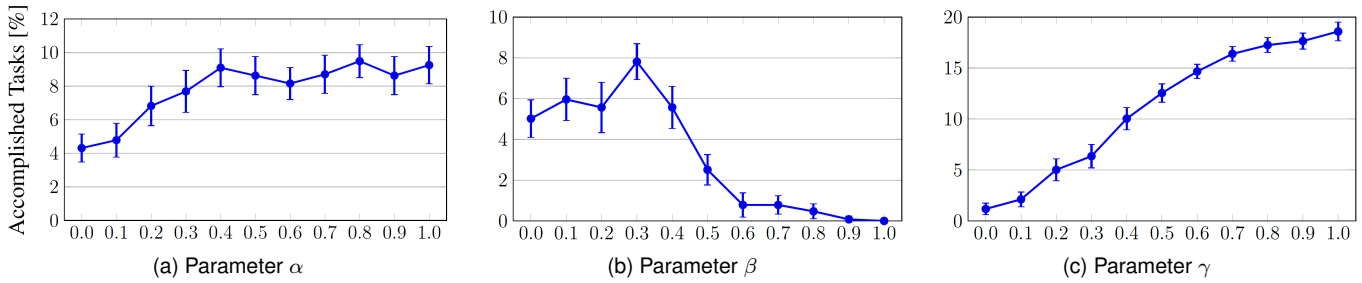


Fig. 14. Impact of the recruiting factor R control parameters on the system performance

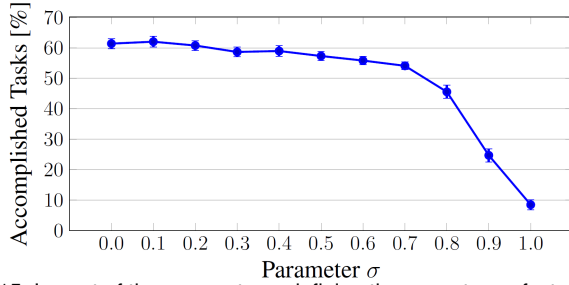


Fig. 15. Impact of the parameter σ defining the acceptance factor A on the system performance

the highest percentage of accomplished tasks. It is worth noting that high values of β , i.e., $\beta > 0.5$ should be avoided. Contacting users for recruitment on the sole basis of their sociability value without considering their location or the remaining charge of the battery lowers the probability to recruit a sufficient number of users. Fig. 14(c) shows the impact of γ , which defines the importance of the energy. The plot shows that the higher the values γ assume, the higher is the number of accomplished tasks. From the analysis, β appears to be the most important for proper setting of the parameters.

Fig. 15 analyzes the impact of the parameter σ , having fixed $\alpha = 0.35$, $\beta = 0.3$ and $\gamma = 0.35$ according to the previous study. The parameter σ defines the importance of sociability and energy for task acceptance, see (8). The majority of the tasks are carried out successfully when user acceptance is mainly driven by energy, i.e., $\sigma < 0.3$. The motivation for the high performance lies in the current evaluation settings, as most of the users have high battery charge at disposal. With the setting $\sigma > 0.8$, the performance dramatically decrease because only highly sociable users accept the task.

7 CONCLUSION

In this paper we have proposed a novel framework for data acquisition in MCS deployed over a fog computing platform. The cloudlets in the fog facilitate user recruitment and task completion. We present a new user recruitment policy called DSE (Distance, Sociability, Energy), which is based on the distance between users and tasks, the user sociability and the energy of user devices. Novel performance metrics are then introduced to assess the efficiency of recruitment policy and the accuracy of task completion. Performance evaluation is conducted with CrowdSenSim simulator in a realistic urban environment for a large number of participants. The results reveal that average number of recruited users improves by nearly 20% if compared to policies using spatial distance

as the only selection criterion. Moreover, the users spend most of the energy for delivering data and not for sensing operation.

As future work, we plan to extend the work in two main directions. On one hand, we will strengthen the simulator capabilities to take into account multiple network technologies as well as a more detailed model of computing, memory and storage requirements of cloudlets. On the other hand, we plan develop a prototype to investigate in realist scenario the performance of the framework in the fog domain.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation (NSF) under Grant No. CNS1464273, and from the National Research Fund, Luxembourg, in the framework of ECO-CLOUD and iSHOP projects.

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] H. B. Dulal and S. Akbar, "Greenhouse gas emission reduction options for cities: Finding the "coincidence of agendas" between local priorities and climate change mitigation objectives," *Habitat International*, vol. 38, pp. 100 – 105, 2013.
- [3] G. Cacciatore, C. Fiandrino, D. Kliazovich, P. Bouvry, and F. Granelli, "Cost analysis of smart lighting solutions for smart cities," in *IEEE ICC*, May 2017.
- [4] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in europe," *Journal of urban technology*, vol. 18, no. 2, pp. 65–82, 2011.
- [5] 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, Feb 2014.
- [6] 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.
- [7] A. D. Cartier, D. H. Lee, B. Kantarci, and L. Foschini, "IoT-big data software ecosystems for smart cities sensing: Challenges, open issues, and emerging solutions," in *International Workshop on Cloud for IoT (CIIoT)*, 2016.
- [8] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Communications Surveys Tutorials*, vol. 17, no. 4, pp. 2347–2376, Fourth 2015.
- [9] F. Bonomi, R. Milito, P. Natarajan, and J. Zhu, "Fog computing: A platform for Internet of Things and analytics," in *Big Data and Internet of Things: A Roadmap for Smart Environments*. Springer, 2014, pp. 169–186.
- [10] F. Bonomi, R. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Workshop on Mobile Cloud Computing*, ser. MCC '12. ACM, 2012, pp. 13–16.
- [11] A. Sciarone, C. Fiandrino, I. Bisio, F. Lavagetto, D. Kliazovich, and P. Bouvry, "Smart probabilistic fingerprinting for indoor localization over fog computing platforms," in *5th IEEE International Conference on Cloud Networking (CloudNet)*, Oct 2016, pp. 39–44.

- [12] G. Cardone, A. Cirri, A. Corradi, L. Foschini, R. Ianniello, and R. Montanari, "Crowdsensing in urban areas for city-scale mass gathering management: Geofencing and activity recognition," *IEEE Sensors Journal*, vol. 14, no. 12, pp. 4185–4195, Dec 2014.
- [13] A. Antonić, V. Bilas, M. Marjanović, M. Matijašević, D. Oletić, M. Pavelić, I. P. Žarko, K. Pripuzić, and L. Skorin-Kapov, "Urban crowd sensing demonstrator: Sense the Zagreb air," in *International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*, Sept 2014, pp. 423–424.
- [14] R. Ganti, F. Ye, and H. Lei, "Mobile crowdsensing: current state and future challenges," *IEEE Comm. Mag.*, vol. 49/11, pp. 32–39, Nov 2011.
- [15] 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, 2013.
- [16] Google Inc., "Google science journal," 2016. [Online]. Available: <https://makingscience.withgoogle.com/science-journal/>
- [17] K. Noyen, D. Volland, D. Wörner, and E. Fleisch, "When money learns to fly: Towards sensing as a service applications using bitcoin," *CoRR*, vol. abs/1409.5841, 2014.
- [18] CEET, "The power of wireless cloud," 2013, White Paper. [Online]. Available: <http://www.ceet.unimelb.edu.au/publications/downloads/ceet-white-paper-wireless-cloud.pdf>
- [19] 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 *Conference on Embedded Networked Sensor Systems*. ACM, 2013, pp. 1–14.
- [20] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "A cost-effective distributed framework for data collection in cloud-based mobile crowd sensing architectures," *IEEE Trans. on Sustainable Computing*, vol. 2, no. 1, pp. 3–16, Jan 2017.
- [21] 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.
- [22] I. Schweizer, R. Bärtil, A. Schulz, F. Probst, and M. Mühlhäuser, "Noisemap - real-time participatory noise maps," in *Workshop on Sensing Applications on Mobile Phones*, 2011, pp. 1–5.
- [23] 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.
- [24] 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 Trans. Sys., Man, and Cybernetics: Systems*, vol. 45/12, pp. 1549–1563, 2015.
- [25] J. Liu, L. Bic, H. Gong, and S. Zhan, "Data collection for mobile crowdsensing in the presence of selfishness," *Journal on Wireless Communications and Networking*, vol. 2016, no. 1, pp. 1–10, 2016.
- [26] J. Li, Y. Zhu, and J. Yu, "Load balance vs utility maximization in mobile crowd sensing: A distributed approach," in *IEEE GLOBECOM*, Austin, USA, December 2014, pp. 259–264.
- [27] 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 Trans. on Computational Social Systems*, vol. 2/3, pp. 109–123, Sep 2015.
- [28] W. Sherchan, P. P. Jayaraman, S. Krishnaswamy, A. Zaslavsky, S. Loke, and A. Sinha, "Using on-the-move mining for mobile crowdsensing," in *IEEE MDM*, July 2012, pp. 115–124.
- [29] 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.
- [30] M. Pouryazdan, B. Kantarci, T. Soyata, and H. Song, "Anchor-assisted and vote-based trustworthiness assurance in smart city crowdsensing," *IEEE Access*, vol. 4, pp. 529–541, 2016.
- [31] M. Pouryazdan and B. Kantarci, "The smart citizen factor in trustworthy smart city crowdsensing," *IEEE IT Professional*, vol. 18/4, pp. 26–33, 2016.
- [32] M. Pouryazdan, C. Fiandrino, B. Kantarci, D. Kliazovich, T. Soyata, and P. Bouvry, "Game-theoretic recruitment of sensing service providers for trustworthy cloud-centric Internet-of-Things (IoT) applications," in *IEEE Globecom Workshops (GC Wkshps)*, Dec 2016, pp. 1–6.
- [33] L. Huang, S. Wang, C.-H. Hsu, J. Zhang, and F. Yang, "Using reputation measurement to defend mobile social networks against malicious feedback ratings," *The Journal of Supercomputing*, vol. 71, no. 6, pp. 2190–2203, 2015.
- [34] S. Reddy, D. Estrin, and M. Srivastava, "Recruitment framework for participatory sensing data collections," in *Springer Pervasive Computing*, 2010, pp. 138–155.
- [35] M. Karaliopoulos, O. Telelis, and I. Koutsopoulos, "User recruitment for mobile crowdsensing over opportunistic networks," in *IEEE INFOCOM*, April 2015, pp. 2254–2262.
- [36] H. Li, T. Li, and Y. Wang, "Dynamic participant recruitment of mobile crowd sensing for heterogeneous sensing tasks," in *IEEE MASS*, Oct 2015, pp. 136–144.
- [37] D. Zhang, H. Xiong, L. Wang, and G. Chen, "Crowdrecruiter: Selecting participants for piggyback crowdsensing under probabilistic coverage constraint," in *Intl. Joint Conf. on Pervasive and Ubiquitous Computing*, 2014, pp. 703–714.
- [38] A. Hassani, P. D. Haghghi, and P. P. Jayaraman, "Context-aware recruitment scheme for opportunistic mobile crowdsensing," in *IEEE ICPADS*, Dec 2015, pp. 266–273.
- [39] C. H. Liu, B. Zhang, X. Su, J. Ma, W. Wang, and K. K. Leung, "Energy-aware participant selection for smartphone-enabled mobile crowd sensing," *IEEE Systems Journal*, pp. 1–12, 2015.
- [40] G. S. Tuncay, G. Benincasa, and A. Helmy, "Participant recruitment and data collection framework for opportunistic sensing: A comparative analysis," in *MobiCom Workshop on Challenged Networks*, ser. CHANTS. ACM, 2013, pp. 25–30.
- [41] L. Pournajaf, L. Xiong, V. Sunderam, and S. Goryczka, "Spatial task assignment for crowd sensing with cloaked locations," in *IEEE MDM*, vol. 1, July 2014, pp. 73–82.
- [42] J. An, X. Gui, Z. Wang, J. Yang, and X. He, "A crowdsourcing assignment model based on mobile crowd sensing in the Internet of Things," *IEEE Internet of Things Journal*, vol. 2, no. 5, pp. 358–369, Oct 2015.
- [43] M. Chen, Y. Hao, Y. Li, C. F. Lai, and D. Wu, "On the computation offloading at ad hoc cloudlet: architecture and service modes," *IEEE Communications Magazine*, vol. 53, no. 6, pp. 18–24, June 2015.
- [44] B. Tang, Z. Chen, G. Hefferman, T. Wei, H. He, and Q. Yang, "A hierarchical distributed fog computing architecture for big data analysis in smart cities," in *ASE BigData & SocialInformatics*. ACM, 2015, pp. 28:1–28:6.
- [45] J. Hoydis, M. Kobayashi, and M. Debbah, "Green small-cell networks," *IEEE Vehicular Technology Magazine*, vol. 6, no. 1, pp. 37–43, March 2011.
- [46] Google Inc., "Google transit," 2016. [Online]. Available: <http://www.google.com/landing/transit/cities/index.html>
- [47] K. K. Jahromi, M. Zignani, S. Gaito, and G. P. Rossi, "Simulating human mobility patterns in urban areas," *Simulation Modelling Practice and Theory*, vol. 62, pp. 137 – 156, 2016.
- [48] X. Pang, W. Benshuai, L. Huifang, and L. Weiwei, "MR-LDA: An efficient topic model for classification of short text in big social data," *International Journal of Grid and High Performance Computing*, vol. 8, no. 4, pp. 100–113, 2016.
- [49] H. Kim, T. Lee, and H.-J. Kim, "A parallel tag affinity computation for social tagging systems using mapreduce," *IJBDDI*, vol. 1, pp. 141–150, 2014.
- [50] N. Haderer, V. Primault, P. Raveneau, C. Ribeiro, R. Rouvoy, and S. Ben Mokhtar, "Towards a practical deployment of privacy-preserving crowd-sensing tasks," in *Middleware Posters & Demos Session*. ACM, 2014, pp. 43–44.
- [51] S. Luo, Y. Sun, Z. Wen, and Y. Ji, "C2: Truthful incentive mechanism for multiple cooperative tasks in mobile cloud," in *IEEE ICC*, May 2016, pp. 1–6.
- [52] S. Wang, L. Huang, C.-H. Hsu, and F. Yang, "Collaboration reputation for trustworthy web service selection in social networks," *Journal of Computer and System Sciences*, vol. 82, no. 1, pp. 130 – 143, 2016.
- [53] J. J. Jussila, H. Kärkkäinen, and J. Multasuo, "Social media roles in crowdsourcing innovation tasks in B2B-relationships," in *ISPIM*, 2013, pp. 1–12.
- [54] C. Robusto, "The cosine-haversine formula," *The American Mathematical Monthly*, vol. 64, no. 1, pp. 38–40, 1957.

- [55] F. Anjomshoa, M. Catalfamo, D. Hecker, N. Helgeland, A. Rasch, B. Kantarci, M. Erol-Kantarci, and S. Schuckers, "Mobile behavior-metric framework for sociability assessment and identification of smartphone users," in *IEEE ISCC*, Jun 2016, pp. 1084–1089.
- [56] S. Scellato, A. Noulas, and C. Mascolo, "Exploiting place features in link prediction on location-based social networks," in *International Conference on Knowledge Discovery and Data Mining*, ser. KDD, 2011, pp. 1046–1054.
- [57] 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 *Intl. Conf. on Mobile Systems, Applications, and Services*, ser. MobiSys. ACM, 2012, pp. 225–238.
- [58] C. Fiandrino, D. Kliazovich, P. Bouvry, and A. Zomaya, "Network-assisted offloading for mobile cloud applications," in *IEEE ICC*, June 2015, pp. 5833–5838.
- [59] 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.
- [60] D. Mendez, M. Labrador, and K. Ramachandran, "Data interpolation for participatory sensing systems," *Pervasive & Mobile Computing*, vol. 9/1, pp. 132–148, 2013.
- [61] D. Hristova, M. J. Williams, M. Musolesi, P. Panzarasa, and C. Mascolo, "Measuring urban social diversity using interconnected geo-social networks," in *International World Wide Web Conference (WWW)*, April 2016.
- [62] F. Anjomshoa, B. Kantarci, M. Erol-Kantarci, and S. Schuckers, "A mobile platform for sociability-based continuous identification," in *IEEE CAMAD*, Oct 2016, pp. 149–151.
- [63] C. Fiandrino, A. Capponi, G. Cacciato, 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. 5, pp. 3490–3503, Feb 2017.
- [64] "FXOS8700CQ: Digital Sensor - 3D Accelerometer + 3D Magnetometer," http://cache.nxp.com/files/sensors/doc/data_sheet/FXOS8700CQ.pdf?pspl=1, 2015.
- [65] "BMP280, Barometric Pressure Sensors," https://www.bosch-sensortec.com/bst/products/all_products/bmp280, 2015.



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 in 2016. 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 IEEE CloudNet 2014 and as TPC member in several IEEE and ACM conferences and workshops. His primary research interests include mobile crowdsensing, mobile cloud/fog computing, and data center communication systems.



Fazel Anjomshoa is a graduate student at the Department of Electrical and Computer Engineering, Clarkson University, Potsdam, NY. He received his M.Sc. and B.Eng. degrees in computer science from Universiti Teknologi Malaysia (UTM) and software engineering from Shahid Bahonar University of Kerman in 2014 and 2011, respectively. He got the best postgraduate student award from UTM for his outstanding contribution during his master studies. His main research interests are big data, cloud computing, internet

of things and distributed systems.



Burak Kantarci (S'05–M'09–SM'12) is an Assistant Professor with the School of Electrical Engineering and Computer Science at the University of Ottawa (Ontario, Canada). From 2014 to 2016, he was an Assistant Professor at the ECE Department at Clarkson University, where he currently holds a courtesy appointment as an assistant professor. Dr. Kantarci received the M.Sc. and Ph.D. degrees in computer engineering from Istanbul Technical University, in 2005 and 2009, respectively. He received the Siemens Excellence Award in 2005 for his studies in optical burst switching. During his Ph.D. study, he studied as a Visiting Scholar with the University of Ottawa, where he completed the major content of his thesis. He has co-authored over 100 papers in established journals and conferences, and contributed to 12 book chapters. He is an Editor of the IEEE Communications Surveys and Tutorials, and an area editor of the IEEE Transactions on Green Communications and Networking. He serves as the Secretary of the IEEE ComSoc Communication Systems Integration and Modeling Technical Committee. He is a member of the ACM and a senior member of the IEEE.



Dzmityr 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). 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. 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.



Jeanna Neefe Matthews is an associate professor of Computer Science at Clarkson University (Potsdam, New York). She is currently a member of the Executive Committee of ACM as the Chair of the Special Interest Group Governing Board and an ACM Distinguished Speaker. Jeanna received her Ph.D. in Computer Science from the University of California at Berkeley in 1999. Her research interests include virtualization, cloud computing, computer security, computer networks and operating systems.