Profiling Energy Efficiency of Mobile Crowdsensing Data Collection Frameworks for Smart City Applications

Mattia Tomasoni[♯], Andrea Capponi[†], Claudio Fiandrino[⋆], Dzmitry Kliazovich[♦], Fabrizio Granelli[♯], and Pascal Bouvry[†]

[♯] Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento, Italy,
[†] FSTC-CSC, University of Luxembourg, Luxembourg, [⋆] IMDEA Networks Institute, Madrid, Spain, [♦] ExaMotive, Luxembourg

E-mails: [♯]mattia.tomasoni@studenti.unitn.it, fabrizio.granelli@unitn.it, [†]{firstname.lastname}@uni.lu,
[⋆]claudio.fiandrino@imdea.org, [♦]kliazovich@ieee.org

Abstract—**Mobile crowdsensing (MCS) has emerged in the last years and has become one of the most prominent paradigms for urban sensing. In MCS, citizens actively participate in the sensing process by contributing data with their smartphones, tablets, wearables and other mobile devices to a collector. As citizens sustain costs while contributing data, i.e., the energy spent from the batteries for sensing and reporting, devising energy efficient data collection frameworks (DCFs) is essential. In this work, we compare the energy efficiency of several DCFs through CrowdSenSim, which allows to perform large-scale simulation experiments in realistic urban environments. Specifically, the DCFs under analysis differ one with each other by the data reporting mechanism implemented and the signaling between users and the collector needed for sensing and reporting decisions. Results reveal that the key criterion differentiating DCFs' energy consumption is the data reporting mechanism. In principle, continuous reporting to the collector should be more energy consuming than probabilistic reporting. However, DCFs with continuous reporting that implement mechanisms to block sensing and data delivery after a certain amount of contribution are more effective in harvesting data from the crowd.**

I. Introduction

The unprecedented growth of the population in urban environments requires rational and sustainable urban development. Smart cities aim at filling this gap by providing the citizens with high-quality services through efficient and rational use of ICT technologies, such as the Internet of Things (IoT) [1]. Sensing is an essential enabler for monitoring infrastructures, transportation systems, environment and health. In this context, including humans in the loop of sensing through their mobile devices has revealed a win-win strategy of mobile crowdsensing (MCS) paradigm [2]. MCS leverages the fusion between complementary roles of human intelligence and mobility with machine intelligence, computational and communication capabilities [3]. Active participation of citizens[1] is one of the key factors of MCS, as it provides a deeper context awareness and a higher coverage compared to traditional sensor networks with no need of further investments [4]. In addition, smart mobile devices act as sensor and communication nodes that are periodically recharged and maintained by their owners.

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

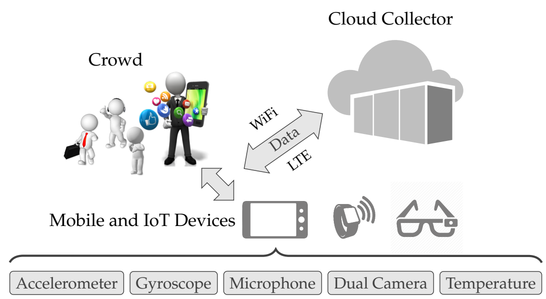


Fig. 1. Cloud-based MCS system

Accelerometer, gyroscope, magnetometer, GPS, microphone and camera are a representative set of sensors embedded in typical smart devices. Mobile devices contribute data to a central collector where it is stored and available to the organizer of a sensing campaign, such as a government agency, an academic institution or a business corporation. The collector is typically located in the cloud and provides shared services and resources to store, analyze and process the received data (see Fig. 1). MCS systems are currently employed to operate applications in health care, environmental and traffic monitoring, management and other domains [5]. To illustrate with a few examples, HazeWatch [6] relies on active citizen participation to monitor air pollution and is currently employed by the National Environment Agency of Singapore on a daily basis. Creekwatch [7] is an application for smartphones developed by the IBM Almaden research center. It allows the monitoring of the conditions of watershed through crowdsensed collected data about the amount of water in the river bed, the amount of trash in the river bank, the flow rate, and a picture of the waterway. Garbage Watch [8] employs citizens to monitor the content of recycling bins with the objective of improving the recycling program.

The organizer is usually responsible for user recruitment and task assignment prior to the start of the sensing campaign and for data analysis and processing [9]. The organization of a MCS campaign requires to sustain costs to reward individuals for their involvement and to verify the accomplishment of the tasks [10] . Consequently, it is crucial to investigate how to maximize the efficiency of a data collection framework (DCF), which is defined in terms of the costs sustained by the organizer and the revenues [11]. MCS follows a Sensing as a Service (S^2aaS) business model, which makes data collected from sensors available to cloud users [12]. Consequently, companies and organizations have no longer the need to acquire an infrastructure to perform a sensing campaign, but they can exploit existing ones recruiting and compensating users for their involvement [13]. The users sustain costs while contributing data too. They spend energy from the batteries for sensing and reporting data and, eventually, consume data subscription plan if cellular connectivity is used for reporting. Developing efficient DCFs is crucial to regulate the degree of user involvement to prevent excessive battery drain from the mobile devices. This is a fundamental limiting factor to foster user participation and contribution [14]. At the same time, the DCFs have to gather a sufficient amount of data to ensure quality of sensed information [15], [16].

The objective of this paper is to analyze and compare the performance of multiple DCFs. To the best of our knowledge, no existing studies have so far studied and compared the amount of data collected and the associated energy costs of several DCFs for large scale sensing campaigns with thousands of users over multiple days. In this work, we consider three DCFs that represent three different families of methodologies. Specifically, we compare the effectiveness of DCFs that differ by the following features: (i) the type of data reporting mechanism implemented, e.g., continuous with stopping mechanisms that prevents users to contribute additional data upon meeting certain criteria or probabilistic, i.e., transmission of sensed data is occasional, and (ii) the degree of control the collector establishes through feedback on the amount of data is still to be harvested. The contribution from a large number of participants is essential to guarantee effectiveness of MCS applications, but prevents researchers to easily perform feasible experiments on real testbeds. Hence, simulations are an excellent alternative and viable solution. CrowdSenSim [17], the first simulator for MCS systems, was designed to fill this gap by providing the researchers a tool to perform large scale simulations over realistic urban environments. For example, its effectiveness has been demonstrated to evaluate performance of city-wide solutions for public street lighting [18].

Our main findings are as follows:

- The data reporting mechanism is the key criterion that differentiates the DCFs. DCFs with continuous reporting that implement mechanisms to block sensing and data delivery based on history of user contribution are more effective in harvesting data from the crowd.
- DCFs with probabilistic reporting exhibit high variability of energy consumption, i.e., to produce the same amount of data, the associated energy cost of different users can be significantly different.
- Human mobility does not influence the behavior of the DCFs. Experiments performed on cities with different urban morphology show that the variation of the average per-user energy consumption achieved with the various DCFs is minor.

II. Background and Motivation

A DCF defines the steps required to acquire data from sensing devices and to perform delivery to the cloud collector. This section overviews existing frameworks for data collection by describing their main characteristics and presents the main challenges in this field.

DCFs are developed to support data collection useful to many applications at the same time. These DCFs usually aim to maximize a set of parameters, e.g., the amount and temporal/spatial coverage of the contributed data or Quality of Information (QoI). At the same time, the DCFs aim to minimize the costs, such as energy consumption or monetary rewards [19]. Wu et al. [20] investigate the most typical trade-off in a DCF between amount of acquired data and energy consumption. Their model analyzes both off-line and on-line settings. In the off-line case, the entire task information is known a-priori and does not change over time. While in the on-line scenario, tasks are allocated dynamically in real-time without any information in advance. First, they provide an optimal algorithm for the off-line setting. Then, they investigate the on-line setting where requests arrive dynamically without prior information, proposing a first-in-first-out (FIFO) task model and an arbitrary deadline (AD) task model. Wang et al. [21] investigate the problem of scheduling several sensing tasks assigned to a user, aiming at ensuring the quality of sensed data while minimizing the energy consumption. Starting from basic cases in which sensing process requires data from only one sensor, the authors define the Minimum Energy Single-sensor task Scheduling (MESS) problem and design a polynomial-time optimal algorithm. Then, they consider a generic case in which sensing tasks need the use of multiple sensors to be accomplished. To solve the problem of Minimum Energy Multi-sensor task Scheduling (MEMS), they propose an integer linear programming (ILP) formulation as well as two effective polynomial-time heuristic algorithms. In [22], the authors propose a fair energy efficient allocation framework whose objective is to minimize the maximum aggregated sensing time. The problem is NP-Hard also when tasks are known a-priori and allocation still has to be done. They firstly investigate the off-line allocation model and propose an efficient polynomial-time approximation algorithm with a factor of $2 - 1/m$, where m is the number of mobile devices joining the system. Then, focusing on the on-line allocation model, they design a greedy algorithm that achieves a ratio of at most m. Han et al. [23] propose an on-line learning algorithm, where a central authority assigns tasks aiming at rewarding participants with a limited amount of budget. It supposes a fixed minimum number of users who actively join the sensing process, while the quality of collected data may vary. Liu et al. [24] propose a method to efficiently select users for participatory crowdsensing. Contributors are dynamically chosen considering their willingness to acquire data and their potential, which is calculated considering the remaining battery in their smartphones. Tasks are distributed with the aim to minimize the probability that an individual does not accomplish the assigned task. CARDAP [25] is a DCF which exploits fog computing platforms to enable efficient data analytics performed in a distributed fashion. The fog allows CARDAP to extend and augment functionality of a previously proposed framework called CAROMM [26]. Similarly to CARDAP, the framework proposed in [27] exploits the fog to perform user recruitment based on multiple criteria, including distance of the participants from the location of the sensing task, their remaining battery charge and the user sociability defined in terms of the amount of time and data users exchange through social media. Fernando et al. [28] propose Honeybee, to make available computing resources between users and share their experience for task classification. Wang et al. [29] present an algorithm to report information in an energy efficient way. It classifies users into two groups. In the first category, the target is to minimize the energy consumption while reporting data and the individuals pay for the data they utilize to the operators. In the second group, users aim to minimize the cost of data reporting using communication technologies, such as WiFi or Bluetooth, which are free-of-charge.

To the best of our knowledge, there are no existing studies which have investigated and compared DCFs for large scale sensing campaigns to mimic a real MCS deployment. Our approach takes into account data contribution and its associated energy costs originated from a multitude of users in city-wide scenarios and over multiple days.

III. Methodology for DCFs Energy Profiling

In this work, we consider three DCFs, which represent different families of data acquisition methodologies characterized by properties and features highlighted in the following paragraphs. Other existing DCFs in the literature exhibit minor variations with respect to the chosen ones. The main differences concern the data reporting phase and can be classified as intermediate solutions of these three main families. This section first presents the studied DCF (Subsection III-A), then outlines the methodology to assess the energy consumption (Subsection III-B), and finally presents the salient features of CrowdSenSim, the employed simulator (Subsection III-C).

A. *DCFs under analysis*

This part presents in detail DCFs under analysis. It focuses on their definitions, objectives, strong and weak aspects. In addition, possible domains of interest in which DCFs could be exploited are presented.

DDF - Deterministic Distributed Framework. DDF is a framework for energy efficient data collection in cloud-based MCS systems that we proposed in a former work [30]. It aims at maximizing the utility of the cloud collector in receiving data from certain sensors in a specific region of interest, while minimizing at the same time the energy costs users sustain to sense and deliver information. The central collector periodically sends to mobile devices beacons to advertise the utility in receiving data from specific sensors in a certain area. Then, the sampling decisions are taken in a distributed fashion at each mobile device locally. Sensing and reporting decisions are driven by environmental context, an estimation of the potential utility and cost of doing sensing and reporting, the level of battery and the amount of data already contributed, and several other parameters. Therefore, the mechanism considers the previous history of the users to determine whether to perform next sensing and reporting operations. This enables fairness among users because prevents data collection from users whose level of battery is too low or that have already contributed considerably in the past.

The applicability of DDF spans across multiple scenarios of interests for smart cities, such as real-time monitoring of the environment or intelligent transportation systems. Such application scenarios require continuous data reporting for an up-to-date analysis of the status of the phenomena observed.

PDA - Probabilistic Distributed Algorithm. Montori et al. [31] propose a distributed algorithm based on probabilistic design to acquire data. The algorithm is based on a limited feedback from the central collector and does not require users completing specific tasks, hence it is in line with the spirit of generic-purpose DCF. The objective of this algorithm is to regulate the amount of data contributed from users in a certain region of interest to avoid data redundancy and energy waste. Additionally, the algorithm aims at providing fairness to the users. Assuming that it is impossible to compute the number of participants in a region of interest because their position is not tracked, the coordinator estimates the required number of participants by computing the number of observations already acquired. The central platform is responsible to set a total per-zone number of observations required to reach a certain level of accuracy in observing a given phenomenon. To reach this goal, the mobile devices decide independently from the central authority whether to perform sampling and reporting. The framework is memoryless because users contribute data independently from the level of previous participation. The range of scenarios where PDA is applicable falls into the same category of DDF.

PCS - Piggyback CrowdSensing. PCS [32] is a DCF that aims at reducing at the minimum any energy cost to promote user participation. The collector does not provide any form of coordination to trigger sensing decisions. Data reporting occurs during the so-called smartphones' opportunities, i.e., sensed data is piggybacked during phone calls or when connected-applications exchange data with remote servers. During these opportunities, the overhead of performing data reporting is low because mobile devices do not have to wake up the radio interface to transmit the collected data on purpose. All the aforementioned features makes PCS suited for delay-tolerant MCS tasks that do not need data to be sent to the central collector in real-time. For instance, PCS could be exploited for mapping non-real-time phenomena like air quality or noise monitoring, requiring only time and place labels or check-ins in mobile social networks.

B. *Proposed methodology for energy profiling*

The energy consumption model takes separately into account both costs associated to reporting and sensing [30]. We denote

(a) Luxembourg City (b) Turin (c) Washington DC

Fig. 2. The set of selected cities. The green dots represent the pedestrian lanes where users walk.

with E_s^c the energy consumption attributed to sensing and with E_s^r the energy cost associated to reporting for sensor s. Then:

$$E_s = E_s^c + E_s^r. \tag{1}$$

The contribution E_s^c due to sensing operation is considered only if sensor s is not already in use for personal usage or another application. Data reporting implies sending data generated from the set of sensors $\mathcal{S}$ to the central collector using available communication technologies. Data transmission is always executed at the beginning of the timeslot t for samples generated during timeslot $t-1$ through WiFi interfaces. WiFi is typically preferred for data delivery in most of the operating systems for mobile devices, including Android and iOS. From an energy perspective, WiFi is more efficient than cellular connectivity (e.g. 3G/LTE) and users do not spend any amount of the monthly data plan from the mobile operators. Assuming that the acquired information is sent to the cloud via WiFi, the energy consumption E^W corresponding to the transmission time T_{tx} is defined as:

$$E^W = \int_0^{T_{tx}} P_{tx}^W \, \mathrm{d}t, \tag{2}$$

where P_{tx}^W is the total device power consumption for transmissions of WiFi packets generated at rate λ_g [33]:

$$P_{tx}^W = \rho_{id} + \rho_{tx} \cdot \tau_{tx} + \gamma_{xg} \cdot \lambda_g. \tag{3}$$

The parameters ρ_{id}, ρ_{tx}, τ_{tx} and γ_{xg} represent respectively the power in idle mode, the power for transmission, the airtime percentage and the energy consumption associated to the processing of each contributed packet.

C. *CrowdSenSim*

We employ CrowdSenSim to perform extensive simulations for the analysis of the energy efficiency of the various DCFs [17]. The simulator models pedestrian mobility in realistic urban environments. To be scalable, the entire walking trajectory of all contributing and non-contributing users in the campaign is known in advance. Hence, during runtime, CrowdSenSim only computes the amount of generated data per user and its associated energy costs due to sensing and reporting [30]. For this work, PDA and PCS DCFs have been implemented to be compared with DDF. The implementation of the DCFs follows the energy consumption model presented in previous subsections. All the practical details about the computation of the energy costs are presented later in Subsection IV-A. Human mobility is defined in the spirit of the ParticipAct dataset originated from a MCS campaign of around 170 students in the Emilia Romagna region (Italy) [34]. Without having at disposal the dataset, we extracted the profile of the average number of contacts during 7 days and used as a reference to determine the user arrivals pattern in CrowdSenSim. Specifically, given the total simulation period in days, we subdivide it into hours and we estimate the minimum number of users to be allocated so that the average user contact follows the ParticipAct profile. A unique user contact is defined as the overlap of two user trajectories within a timeslot, i.e., their distance is below a given radius R. Note that multiple overlaps in one timeslot still count as unique contacts, while multiple overlaps in different timeslots count as separated contacts.

IV. Performance Evaluation

Performance evaluation exploits CrowdSenSim [17]. This section first presents its detailed features, then the simulation scenario and the output results.

A. *Simulation Set-up*

In CrowdSenSim the layout of any city consists of a set of coordinates $\mathcal{C}$ containing information on *<latitude, longitude, altitude>* that defines the streets of the city. $\mathcal{C}$ is obtained with a procedure. At first, CrowdSenSim downloads the walkable city graph of OpenStreetMap (OSM) through OSMnx Python package [35]. Unfortunately, OSM street nodes are inconsistent for direct use in CrowdSenSim because they include dead-ends, intersections and all the points in a segment when the streets curve. OSMnx automatically simplifies and corrects the street topology through an algorithm by removing those points and unifying each resulting set of sub-edges into single edges. However, the resulting topology still lacks of a sufficiently fine-grained level of detail. Hence, CrowdSenSim runs in the background an algorithm that augments the precision of the OSMnx topology by adding nodes on the streets with user-defined level of detail, for example 1 m.

For the experiments, the cities of Luxembourg, Turin (Italy) and Washington DC (USA) were selected. The rational of the choice is twofold. First, the objective is to consider cities growing in size. The center of Luxembourg City covers an area of $51.47\ \mathrm{km}^2$ with a population of 114 090 inhabitants as of the

TABLE I
SENSOR AND COMMUNICATION EQUIPMENT PARAMETERS USED FOR PERFORMANCE EVALUATION

SENSOR	PARAMETER	VALUE	UNIT
Accelerometer	Sample rate	4	kHz
	Sample size	6	Bytes
	Current	450	µA
Proximity	Sample rate	8.1	MHz
	Sample size	2	Bytes
	Current	150	µA
GPS	Update period	10	s
	Sample size	24	Bytes
	Current	23	mA

(a) Sensor Equipment

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 Equipment

end of 2016 and is the home of many national and international institutional buildings (see Fig. 6(a)). The city center of Turin occupies an area of 130.17 km^2 and has a population of $883\,601$ inhabitants as of the beginning of 2016 (see Fig. 6(b)). The city center of Washington DC covers approximately an area of 158.1 km^2 with a resident population of $672\,228$ inhabitants as of the end of 2015 (see Fig. 6(c)). The second reason for such a choice is the urban morphology, which defines the topology of the street network. Luxembourg City exhibits the typical north european urban morphology with many short streets with small lanes, a high density of crossroads in the center and few parallel large streets in the periphery. Washington DC differs completely from Luxembourg City and its street network topology presents a high number of parallel long streets with large lanes. In addition, the differences between the urban morphology in the city center and the periphery are minimal. Turin falls in between the two former categories because of typical roman grid street organization.

As stated in Section III-C, the user arrival pattern in the system is based on realistic mobility traces for a time period of 12 consecutive hours in one day. The PartecipAct dataset provides information on the user contacts per-hour. Hence, the simulator allocates during each hour a certain amount of users to reach the desired number of contacts. Specifically, we count one user contact if two overlapping trajectories overlap within a timeslot at least once, i.e., the two users are within a distance R of 50 m. Each user has only one device contributing data and walks for a period of time that is uniformly distributed between $[1, 40]$ minutes with an average speed uniformly distributed between $[1, 1.5]$ m/s. The users move in a random walk fashion over the street graph through a random generated starting and arrival point while satisfying the time constraint of the walk. Unless otherwise stated, the number of participants is fixed to $20\,000$. For each user device, the initial battery level at the beginning of the walk is uniformly distributed in the range [80-90]%. The full battery capacity varies according to the model of the phone that is randomly chosen from a list that includes the most popular models currently available on the market: 2200 mAh (Huawei P8 Lite), 2550 mAh (Samsung Galaxy S6), 2800 mAh (LG G5) and 3300 mAh (Samsung Galaxy J7).

Data generation is implemented using heterogeneous sensing equipment commonly available in today mobile devices, including the MPU-6500 3axis linear accelerometer from InvenSense, the TMD4903 proximity sensor from AMS and the SKG13BL GPS module from SKYLAB. Communications occur over the WiFi link, having obtained the precise location of WiFi hotspots in form of <*latitude, longitude*>. In the simulator, the time is slotted. The maximum quantity of data that can be sent in a time slot is defined by the transmission speed, fixed to 1 Mbps. Table I presents the detailed information on communication and the sensor equipment.

The DCFs are implemented as follows. With DDF, the users perform continuous sensing and reporting driven by the collector feedback until they reach a decrease of battery level of 0.5%. Then, they stop contributing data and the associated energy consumption is computed at the granularity of the timeslot, i.e., 1 minute. In PDA, the users continuously generate data and every minute they determine the probability of delivering the acquired data. When no transmissions occur, data is stored locally on a buffer whose occupancy increases and decreases with the number of successful delivery attempts. Finally, PCS implements a buffer mechanism as well to store the acquired data that is delivered during phone calls. The distribution and the duration of daily phone calls follow the profile of weekday 1 in [36], that is computed by normalizing the average call arrival rate and average calls duration within 24 hours from a dataset of four different days.

B. Simulation Results

For performance evaluation, we first evaluate the distribution of the energy consumption for the three DCFs in Luxembourg City. Then, we investigate the performance of the DCFs for various cities and show for a limited number of users the active contribution periods to highlight the differences between the DCFs. Finally, we assess the amount of collected data.

Energy Consumption. Fig. 3 presents the CDF of the per-user battery drain for the considered DCFs in Luxembourg City. By design, DDF includes a stopping mechanism to prevent users contribute additional data upon meeting given criteria, such as if the battery drain attributed to previous sensing and reporting operations has exceeded a given threshold or if the amount of previous contributed data has reached a certain value. Hence, it limits the maximum energy consumption the users

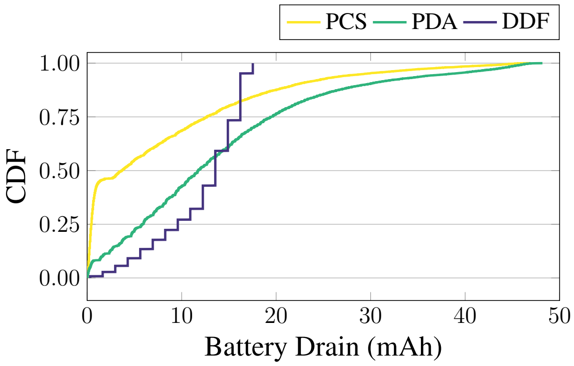


Fig. 3. CDF of battery drain per user for different DCFs in Luxembourg City

spend, i.e., in this experiment all the users spend at maximum 17 mAh. Comparatively, the percentage of users that spend more than 17 mAh is 20% and 30% for with PDA and PCS respectively. Interestingly, DDF lowers the number of users with low energy consumption. This means that the organizer of the sensing campaign effectively exploits the users that agreed to participate and contribute data and that are compensated for such contribution. On the contrary, a significant fraction of users consume a little amount of energy with PCS and PDA. The reason is the probabilistic data delivery mechanism that if applied for periods of time in the order of hours prevents some of the users to transmit significant amounts of data. Note that in the context of crowdsensing, the dominant factor affecting energy consumption is data delivery and not sensing [30].

Fig. 4 shows the CDF of battery drain per user in different cities with the considered DCFs. Interestingly, the various DCFs behave similarly within the same city and the minor variation is attributed to Washington DC. Consequently, the size of the city has a minor impact on the performance of the DCFs. Note that DDF exhibits a CDF that mimics a step function. Each step identifies the group of users that stopped contributing data because of the stopping mechanism and have delivered to the system a similar amount of data.

Fig. 5 shows the amount of collected data and the associated battery drain for all the DCFs. Each mark of the plot represents the energy consumptions that a set of users has spent to produce a given amount of data. First, it should be noted that DDF exhibits a low number of marks. The reason is that the users exhibit a similar behaviour as DDF indirectly controls the level of energy consumption. Viceversa, the other DCFs exhibit much higher variability due to the probabilistic reporting: to produce the amount of data, users spend a different amount of energy. This variability becomes higher as the total amount of data increases. Practically, the result shows that providing user rewards on sole basis of the amount of contributed data fails to properly compensate for users' costs because of the technical implementation of data reporting.

Amount of Collected Data. Fig. 6 shows the trajectories of five users walking in Luxembourg City that contribute data with the various DCFs. The objective is to highlight the active periods of contribution to clearly show the differences between the reporting mechanisms. With DDF, data contribution is continuous until users stop sending data because of the sufficient amount of contribution. With PDA, users generate data in an intermittent fashion depending on the probability. PCS shows that a user can also not contributing in case during the walking period no calls or applications are exploited.

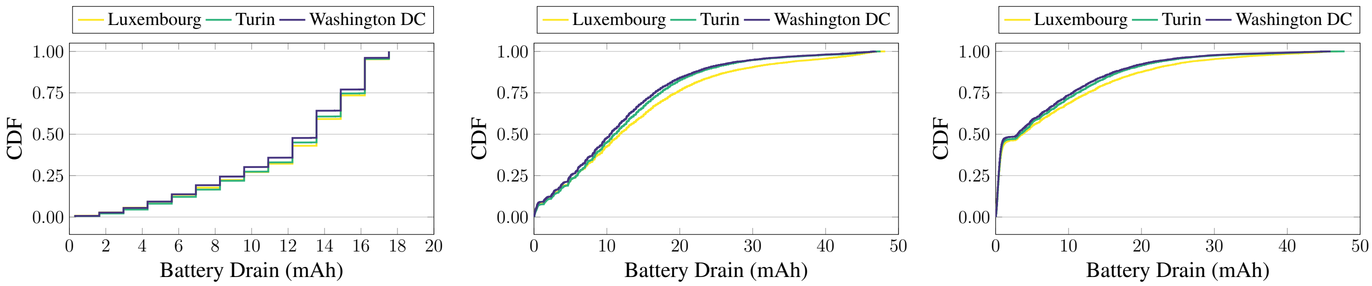
Fig. 7 shows in form of heatmap the spatial distribution of the total amount of collected data at the end of the simulation period for Luxembourg. The heatmap is normalized between 0 and 1, where 1 indicates a total of 10 MB of data generated during the entire simulation period. PDA achieves a higher spatial distribution of amount of collected data than the others DCFs. This is because it does not include any mechanism to stop contribution. DDF shows a lower amount of collected data due to the stopping mechanism, permitting energy savings as shown in Fig. 3. PCS achieves the lowest amount of contributed data. Indeed, although users perform continuous sensing, data reporting fully depends on the probability of performing calls.

Fig. 8 shows the amount of collected data in Luxembourg City comparing the considered DCFs for different number of users. PDA is the DCF that contributes the highest amount of data, as users are not prevented by any stopping mechanisms. DDF presents a big amount of data in the first phases due to continuous reporting, but then users stop to save energy. On the contrary, PCS achieves the lowest amount of data collected and fails to capture area of interests with particular accuracy. Again, the motivation lies in the reporting mechanism implemented.

V. Conclusion

A DCF defines the efficiency of a MCS system in terms of energy consumption and quality of information acquired. Effective frameworks aim to make minimal the energy costs associated to sensing and reporting. Profiling energy is crucial to assess the costs of a sensing campaign and to plan proper user incentives plans like monetary rewarding. In this paper, we evaluated multiple DCFs through large scale simulations in realistic urban scenarios. Our methodology takes into account energy costs due to sensing and reporting processes of each mobile device and scales them in large urban scenarios exploiting the CrowdSenSim simulator. We showed that the data reporting mechanism is the key criterion that differentiates the DCFs and DCFs with probabilistic reporting comparatively achieve higher energy consumption. Furthermore, such DCFs present high variability, meaning that to produce the same amount of data, the associated energy cost of different users can be significantly different. Consequently, DCFs with continuous reporting that implement mechanisms to block sensing and data delivery after a certain amount of contribution are more effective in harvesting data from the crowd. Finally, human mobility does not influence the behavior of the DCF. Experiments performed on cities with different urban morphology show that the average per-user energy consumption achieved with the various DCF exhibits minor variations.

As future work, we plan to extend the current results by developing an application that implements the considered DCFs with the objective to profile their energy consumption by measuring with a power monitor the corresponding current drain. This will allow to obtain realistic results that are more



(a) DDF (b) PDA (c) PCS

Fig. 4. CDF of battery drain per user for considered DCFs in different cities

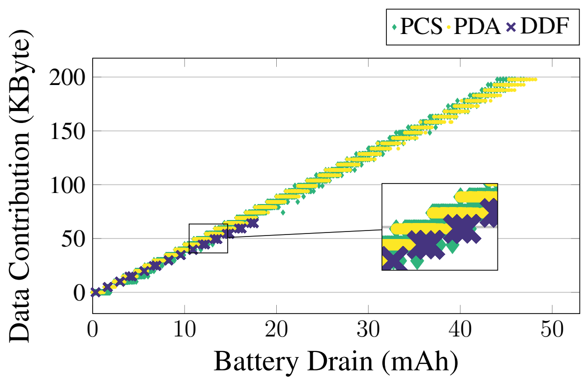


Fig. 5. Amount of collected data and the associated battery drain

accurate than the ones obtained through the model and to compare in a more realistic way the DCFs.

ACKNOWLEDGMENT

The authors would like to thank Piergiorgio Vitello for his feedback on the draft.

REFERENCES

[1] 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.

[2] L. Duan, L. Huang, C. Langbort, A. Pozdnukhov, J. Walrand, and L. Zhang, "Human-in-the-loop mobile networks: A survey of recent advancements," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 4, pp. 813–831, Apr 2017.

[3] 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.

[4] 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.

[5] 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 Quarter 2013.

[6] V. Sivaraman, J. Carrapetta, K. Hu, and B. G. Luxan, "Hazewatch: A participatory sensor system for monitoring air pollution in sydney," in *IEEE Conference on Local Computer Networks (LCN) - Workshops*, Oct 2013, pp. 56–64.

[7] S. Kim, C. Robson, T. Zimmerman, J. Pierce, and E. M. Haber, "Creek Watch: Pairing usefulness and usability for successful citizen science," in *Proc. of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI. ACM, 2011, pp. 2125–2134.

[8] S. Reddy and V. Samanta, "Urban Sensing: Garbage Watch," 2011, UCLA Center for Embedded Networked Sensing.

[9] 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.

[10] M. Pouryazdan, C. Fiandrino, B. Kantarci, T. Soyata, D. Kliazovich, and P. Bouvry, "Intelligent gaming for mobile crowd-sensing participants to acquire trustworthy big data in the internet of things," *IEEE Access*, vol. 5, pp. 22 209–22 223, Oct 2017.

[11] A. Capponi, C. Fiandrino, D. Kliazovich, P. Bouvry, and S. Giordano, "Energy efficient data collection in opportunistic mobile crowdsensing architectures for smart cities," in *IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, May 2017, pp. 307–312.

[12] A. Capponi, C. Fiandrino, C. Franck, U. Sorger, D. Kliazovich, and P. Bouvry, "Assessing performance of internet of things-based mobile crowdsensing systems for sensing as a service applications in smart cities," in *IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, Dec 2016, pp. 456–459.

[13] 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.

[14] C. Fiandrino, B. Kantarci, F. Anjomshoa, D. Kliazovich, P. Bouvry, and J. Matthews, "Sociability-driven user recruitment in mobile crowdsensing internet of things platforms," in *IEEE Global Communications Conference (GLOBECOM)*, Dec 2016, pp. 1–6.

[15] B. Song, H. Shah-Mansouri, and V. W. S. Wong, "Quality of sensing aware budget feasible mechanism for mobile crowdsensing," *IEEE Transactions on Wireless Communications*, vol. 16, no. 6, pp. 3619–3631, June 2017.

[16] F. Restuccia, N. Ghosh, S. Bhattacharjee, S. K. Das, and T. Melodia, "Quality of information in mobile crowdsensing: Survey and research challenges," *CoRR*, vol. abs/1709.03583, 2017.

[17] C. Fiandrino, A. Capponi, G. Cacciatore, D. Kliazovich, U. Sorger, P. Bouvry, B. Kantarci, F. Granelli, and S. Giordano, "CrowdSenSim: a simulation platform for mobile crowdsensing in realistic urban environments," *IEEE Access*, vol. 5, pp. 3490–3503, Feb 2017.

[18] G. Cacciatore, C. Fiandrino, D. Kliazovich, F. Granelli, and P. Bouvry, "Cost analysis of smart lighting solutions for smart cities," in *IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.

[19] Z. Zheng, Y. Peng, F. Wu, S. Tang, and G. Chen, "Trading data in the crowd: Profit-driven data acquisition for mobile crowdsensing," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 2, pp. 486–501, Feb 2017.

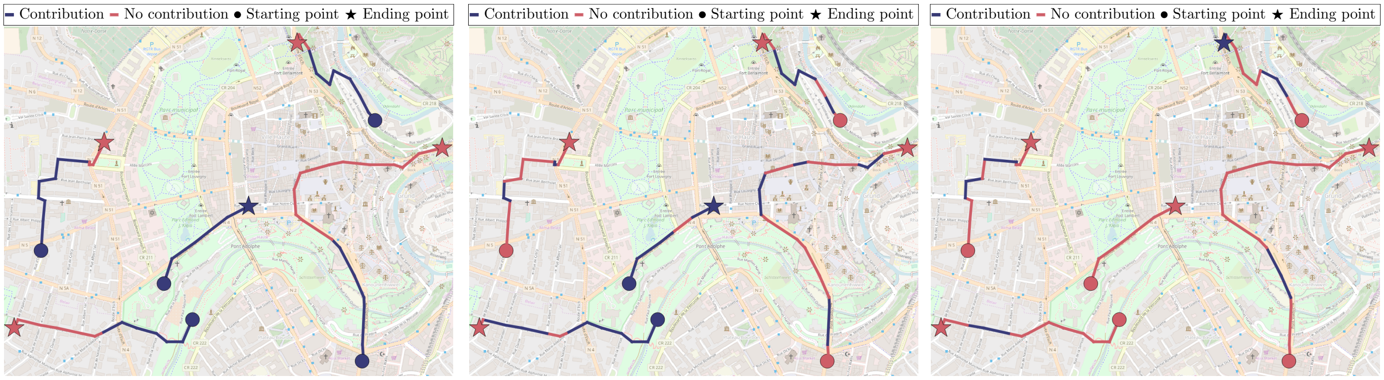
[20] W. Wu, J. Wang, M. Li, K. Liu, F. Shan, and J. Luo, "Energy-efficient transmission with data sharing in participatory sensing systems," *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 12, pp. 4048–4062, Dec 2016.

[21] J. Wang, J. Tang, G. Xue, and D. Yang, "Towards energy-efficient task scheduling on smartphones in mobile crowd sensing systems," *Computer Networks*, vol. 115, pp. 100 – 109, 2017.

[22] Q. Zhao, Y. Zhu, H. Zhu, J. Cao, G. Xue, and B. Li, "Fair energy-efficient sensing task allocation in participatory sensing with smartphones," in *IEEE Conference on Computer Communications (INFOCOM)*, April 2014, pp. 1366–1374.

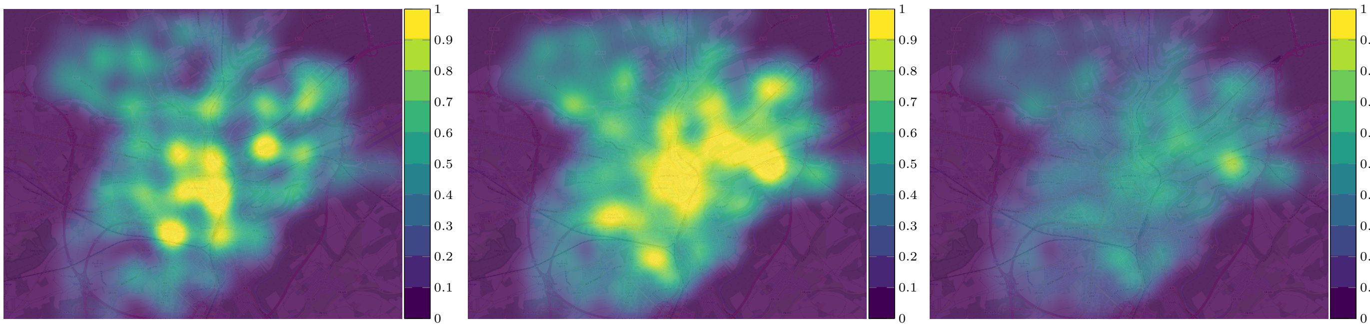
[23] K. Han, C. Zhang, and J. Luo, "Taming the uncertainty: Budget limited robust crowdsensing through online learning," *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1462–1475, June 2016.

[24] 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*, vol. 11, no. 3, pp. 1435–1446, Sep 2017.



(a) DDF (b) PDA (c) PCS

Fig. 6. User trajectories with the associated data contribution in Luxembourg City



(a) DDF (b) PDA (c) PCS

Fig. 7. Normalized distribution of amount of contributed data in Luxembourg City comparing different DCFs

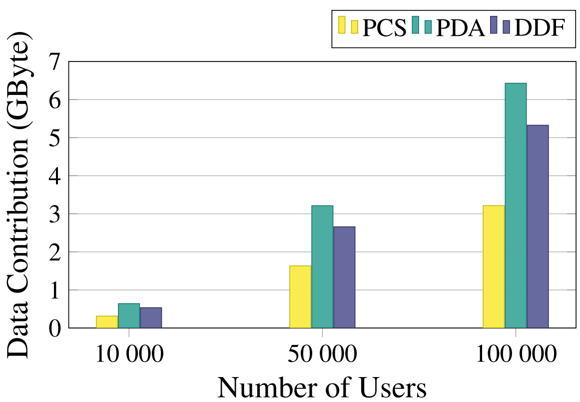


Fig. 8. Amount of contributed data for considered DCFs in Luxembourg City

[25] 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.

[26] 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.

[27] C. Fiandrino, F. Anjomshoa, B. Kantarci, D. Kliazovich, P. Bouvry, and J. N. Matthews, "Sociability-driven framework for data acquisition in mobile crowdsensing over fog computing platforms for smart cities," *IEEE Transactions on Sustainable Computing*, vol. 2, no. 4, pp. 345–358, Oct 2017.

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

[29] 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.

[30] 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 Transactions on Sustainable Computing*, vol. 2, no. 1, pp. 3–16, Jan 2017.

[31] F. Montori, L. Bedogni, and L. Bononi, "Distributed data collection control in opportunistic mobile crowdsensing," in *Proc. of the 3rd Workshop on Experiences with the Design and Implementation of Smart Objects*, ser. SMARTOBJECTS. ACM, 2017, pp. 19–24.

[32] 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.

[33] 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, 2015.

[34] S. Chessa, M. Girolami, L. Foschini, R. Ianniello, A. Corradi, and P. Bellavista, "Mobile crowd sensing management with the ParticipAct living lab," *Pervasive and Mobile Computing*, vol. 38, pp. 200 – 214, 2017.

[35] G. Boeing, "OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks," *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139, 2017.

[36] D. Willkomm, S. Machiraju, J. Bolot, and A. Wolisz, "Primary user behavior in cellular networks and implications for dynamic spectrum access," *IEEE Communications Magazine*, vol. 47, no. 3, pp. 88–95, March 2009.